

SEARCH REQUEST FORM

50597

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Searcher: _____
 Contact: Sheppard

Searcher Phone #: _____
 Tel: 308-4499

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed: 9/12/07

Searcher Prep & Review Time: _____

Clerical Prep Time: _____

Online Time: _____

Type of Search

NA Sequence (#) _____

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbi: _____

Dr. Link _____

Lexis/Nexis _____

Sequence Systems _____

WWW/Internet _____

Other (specify) _____

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:48:30 ; Search time 21.52 Seconds
(without alignments)
387,324 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346

Sequence: 1 ADRAAVPIVNLKDELPPSW.....DVLRHPTWPKQSVHGSDDPN 63

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP:PREMBL_16.*
2: SP:archaea.*
3: SP:bacteria.*
4: SP:fungi.*
5: SP:human.*
6: SP:invertebrate.*
7: SP:mhc.*
8: SP:organelle.*
9: SP:phage.*
10: SP:plant.*
11: SP:rodent.*
12: SP:unclassified.*
13: SP:vertebrate.*
14: SP:virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294	85.0	184	11 09UK63	09UK63 mus musculu
2	294	85.0	1140	11 061434	061434 mus musculu
3	294	85.0	1774	11 062001	062001 mus musculu
4	284	82.1	171	11 09W0W5	09W0W5 rattus norv
5	284	82.1	226	11 09Q2D2	09Q2D2 rattus norv
6	255	73.7	386	13 093419	093419 gallus galli
7	197	56.9	1367	11 035206	035206 mus musculu
8	196	56.6	1367	11 09E0D9	09E0D9 mus musculu
9	193	55.8	1388	4 09Y4W4	09Y4W4 homo sapien
10	169.5	49.0	581	5 09VS09	09VS09 drosophila
11	140	40.5	650	5 017866	017866 caenorhabdi
12	140	40.5	778	5 09U9K6	09U9K6 caenorhabdi
13	140	40.5	1117	5 09U9K7	09U9K7 caenorhabdi
14	70.5	20.4	317	4 043467	043467 homo sapien
15	67.5	19.5	614	4 003989	003989 homo sapien
16	65	18.8	1612	5 09VYQ2	09VYQ2 drosophila
17	64	18.5	644	4 09YA13	09YA13 aeropyrum p
18	63	18.2	336	2 09XG16	09XG16 streptomyce
19	62.5	18.1	467	5 09XX44	09XX44 caenorhabdi

20	62.5	18.1	620	2 031526	031526 bacillus su
21	62	17.9	325	4 09Y247	09Y247 homo sapien
22	62	17.9	419	5 091096	091096 caenorhabdi
23	61.5	17.8	362	7 09MXH0	09MXH0 pan troglod
24	61.5	17.8	1172	14 09E1Y8	09E1Y8 ceropithec
25	61	17.6	738	11 09E213	09E213 rattus norv
26	60.5	17.5	862	10 09FT17	09FT17 lycopersico
27	60	17.3	258	1 P94920	P94920 methanosarc
28	60	17.3	334	11 09WTU8	09WTU8 mus musculu
29	60	17.3	507	2 P95552	P95552 pseudomonas
30	60	17.3	984	2 09XCD4	09XCD4 thermomonas
31	60	17.3	1280	2 09F6X9	09F6X9 chloroflexu
32	59.5	17.2	181	7 09WVC1	09WVC1 homo sapien
33	59.5	17.2	246	2 09RCE1	09RCE1 xanthomonas
34	59.5	17.2	268	2 09RXX2	09RXX2 delnocosus
35	59	17.1	91	7 019588	019588 homo sapien
36	59	17.1	165	7 09MW45	09MW45 homo sapien
37	59	17.1	303	3 074432	074432 schizosacch
38	58.5	16.9	181	7 095432	095432 pan troglod
39	58.5	16.9	305	7 030991	030991 pan troglod
40	58.5	16.9	354	7 095530	095530 pan troglod
41	58.5	16.9	362	7 09TPL5	09TPL5 pan troglod
42	58.5	16.9	362	7 09MXL9	09MXL9 pan troglod
43	58.5	16.9	362	7 09MXK5	09MXK5 pan troglod
44	58.5	16.9	362	7 09MXI2	09MXI2 pan troglod
45	58.5	16.9	362	7 09MXI0	09MXI0 pan troglod

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	184 AA.
09UK63				
AC 09UK63:				
DT 01-OCT-2000 (TREMBLrel. 15, Created)				
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)				
DE ENDOSTATIN (FRAGMENT) :				
OS Mus musculus (Mouse) :				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX NCBI_Taxid=10090;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN=CHINESE KUNMING;				
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;				
RT "Anticancer treatment of targeted fusion protein delivery to tumor				
RT neovasculation."				
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.				
DR EMBL; AF257775; AAF69009.1; -				
FT NON_TER	1			
FT NON_TER	184			
FT NON_TER	184			
SQ SEQUENCE	184 AA;	20376 MW;	AC06F9D8D103412A	CRC64;

Query Match 85.0%; Score 294; DB 11; Length 184;
Best local Similarity 81.0%; Pred. No. 2.6e-28;
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY	1	ADRAAVPIVNLKDELLFPSEALFSGSEGLPKGARIFSFDDGVLRHPTWPKQSVHGS	60
DB	64	ADGASVPIVNLKDEVLSPSWDSLFPSSGQGLPGARIFSFDDGVLRHPTWPKQSVHGS	123
QY	61	DPN 63	
DB	124	DPS 126	

RESULT 2
061434
ID 061434
AC 061434;
PRELIMINARY;
PRT: 1140 AA.

DT	01-NOV-1996	(TREMBLrel. 01, Created)
DT	01-NOV-1996	(TREMBLrel. 01, last sequence update)
DT	01-JUN-2000	(TREMBLrel. 14, last annotation update)
DE	COLLAGEN (FRAGMENT).	
GN	COL15A1.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxId=10090;	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=9405075; PubMed=8240330;	
RA	Abe N., Muragaki Y., Yoshioka H., Inoue H., Niinomiya Y.;	
RT	"Identification of a novel collagen chain represented by extensive	
RT	interruptions in the triple-helical region.";	
RL	Cell. Mol. Biol. Res. 196;576-582(1993).	
DR	EMBL; D17546; BAA04483.1; -.	
DR	HSSP; P39061; IKOE.	
DR	MGD; MGI:88449; COL15a1.	
DR	InterPro; IPR000087; -.	
DR	Pfam; PF01391; Collagen; 6.	
FT	NON_TER	
SQ	SEQUENCE	1140 AA; 11516 MW; 880C76E682B3BDFE CRC64;

FT NON_TER 1 1
 FT NON_TER 171 171
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 171;
 Best Local Similarity 79.4%; Pred. No. 4.1e-27;
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPGSEALFSGSEGLPKPGARIFSGDKDYLRRHPTWPKSVWHS 60
 DB 56 ADRAAPVIVNLKDEVLFPSPMDTLFSGSQGLHSGARIFSGDRDYLRRHPTWPKSVWHS 115
 OY 61 DPN 63
 DB 116 DPS 118

RESULT 5
 090ZD2 PRELIMINARY; PRT; 226 AA.

ID 090ZD2
 AC 090ZD2
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE COLLAGEN XVIII (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RA Chen L., Perletti G., Folkman J.;
 RT "Antitumor activity of rat endostatin."
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF189709; AAF00975.1; -;
 DR HSP; P39061; IKOE.
 KW Collagen.
 FT NON_TER
 SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 226;
 Best Local Similarity 79.4%; Pred. No. 5.6e-27;
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPGSEALFSGSEGLPKPGARIFSGDKDYLRRHPTWPKSVWHS 60
 DB 106 ADRAAPVIVNLKDEVLFPSPMDTLFSGSQGLHSGARIFSGDRDYLRRHPTWPKSVWHS 165
 OY 61 DPN 63
 DB 166 DPS 168

RESULT 6
 093419 PRELIMINARY; PRT; 386 AA.

ID 093419
 AC 093419
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE COLLAGEN XVIII (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxId=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Halfter W., Dong S., Schurer B., Cole G.;
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan."

RL J. Biol. Chem. 0:0-0(1998).
 DR EMBL; AF083440; AAC33294.1; -;
 DR HSP; P39061; IKOE.
 DR InterPro; IPR000087; -;
 FT NON_TER
 SQ SEQUENCE 386 AA; 41775 MW; 34D40FA09EBA3B0E CRC64;

Query Match 73.7%; Score 255; DB 13; Length 386;
 Best Local Similarity 72.1%; Pred. No. 3.6e-23;
 Matches 44; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPGSEALFSGSEGLPKPGARIFSGDKDYLRRHPTWPKSVWHS 60
 DB 266 ADRAAPVIVNLKDEVLFPSPMDTLFSGSQGLHSGARIFSGDRDYLRRHPTWPKSVWHS 325
 OY 61 DPN 61
 DB 326 D 326

RESULT 7
 035206 PRELIMINARY; PRT; 1367 AA.

ID 035206
 AC 035206
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE PROCOLLAGEN, TYPE XV (TYPE XV COLLAGEN).
 GN COL15A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97480713; PubMed=9339358;
 RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
 RA Pihlajaniemi T.;
 RT "Cloning of mouse type XV collagen sequences and mapping of the
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
 RT (XV) collagen sequences indicates divergence in the number of small
 RT collagenous domains."
 RL Genomics 45:31-41(1997).
 DR EMBL; AF011450; AAC5387.1; -;
 DR HSP; P39061; IKOE.
 DR MGD; MGI:88449; Col15a1.
 DR InterPro; IPR000087; -;
 DR InterPro; IPR001791; -;
 DR InterPro; IPR003129; -;
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF02210; TSPN; 1.
 DR SMART; SM00282; LamG; 1.
 SQ SEQUENCE 1367 AA; 140524 MW; A483A1254AF3AECC CRC64;

Query Match 56.9%; Score 197; DB 11; Length 1367;
 Best Local Similarity 52.4%; Pred. No. 1.9e-15;
 Matches 33; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPGSEALFSGSEGLPKPGARIFSGDKDYLRRHPTWPKSVWHS 60
 DB 1247 ADRAAPVIVNLKDEVLFPSPMDTLFSGSQGLHSGARIFSGDRDYLRRHPTWPKSVWHS 1306
 OY 61 DPN 63
 DB 1307 NPH 1309

RESULT 8
 09EOD9 PRELIMINARY; PRT; 1367 AA.
 ID 09EOD9
 AC 09EOD9

ID	Q914M4	PRELIMINARY	PRT	1388	AA
DT	01-MAR-2001 (TREMBLER, 16, Created)				
DT	01-MAR-2001 (TREMBLER, 16, Last annotation update)				
DT	01-MAR-2001 (TREMBLER, 16, Last annotation update)				
DE	TYPE XV COLLAGEN.				
GN	COL15A1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129/SV;				
RX	MEDLINE=97480713; PubMed=9339358;				
RA	Hagg P.M., Horelli-Kutlun N., Eklund L., Palotie A.,				
RT	Phliaglanem T.;				
RT	*Cloning of mouse type XV collagen sequences and mapping of the				
RT	corresponding gene to 4B1-3. Comparison of mouse and human alpha 1				
RT	(XV) collagen sequences indicates divergence in the number of small				
RL	collagenous domains.*;				
RL	Genomics 45:31-41(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129/SV;				
RX	Pubmed=11068203;				
RA	Eklund L., Muona A., Lietaard J., Phliaglanem T.;				
RT	*Structure of the mouse type XV collagen gene, Col15a1, comparison				
RT	with the human COL15A1 gene and functional analysis of the promoters				
RT	of both genes.*;				
RL	Matrix Biol. 19:489-500(2000).				
DR	EMBL: AF261131; AAG27545.1; JOINED.				
DR	EMBL: AF261109; AAG27545.1; JOINED.				
DR	EMBL: AF261110; AAG27545.1; JOINED.				
DR	EMBL: AF261111; AAG27545.1; JOINED.				
DR	EMBL: AF261112; AAG27545.1; JOINED.				
DR	EMBL: AF261113; AAG27545.1; JOINED.				
DR	EMBL: AF261114; AAG27545.1; JOINED.				
DR	EMBL: AF261115; AAG27545.1; JOINED.				
DR	EMBL: AF261116; AAG27545.1; JOINED.				
DR	EMBL: AF261117; AAG27545.1; JOINED.				
DR	EMBL: AF261118; AAG27545.1; JOINED.				
DR	EMBL: AF261119; AAG27545.1; JOINED.				
DR	EMBL: AF261120; AAG27545.1; JOINED.				
DR	EMBL: AF261121; AAG27545.1; JOINED.				
DR	EMBL: AF261122; AAG27545.1; JOINED.				
DR	EMBL: AF261123; AAG27545.1; JOINED.				
DR	EMBL: AF261124; AAG27545.1; JOINED.				
DR	EMBL: AF261125; AAG27545.1; JOINED.				
DR	EMBL: AF261126; AAG27545.1; JOINED.				
DR	EMBL: AF261127; AAG27545.1; JOINED.				
DR	EMBL: AF261128; AAG27545.1; JOINED.				
DR	EMBL: AF261129; AAG27545.1; JOINED.				
DR	EMBL: AF261130; AAG27545.1; JOINED.				
KW	Collagen.				
SO	SEQUENCE. 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;				
Query Match	56.6%; Score 196; DB 11; Length 1367;				
Best Local Similarity	53.2%; Pred. No. 2.5e-15;				
Matches	33; Conservative 13; Mismatches 16; Indels 0; Gaps 0;				
OY	1 ADRAAPVIVNLKDELLPSPWEALFGSGSEGPLKPGARIFSFDEKDLVRHTPWPQKSYWNGS 60				
DB	1247 AERFCLPVLNLTGQVILFNNWDSIFSGDGGQFWTHLPIYSFQDRDVTDPSPWQKVVWNGS 1306				
OY	61 DP 62				
DB	1307 NP 1308				
RESULT	9				
ID	O914M4	PRELIMINARY	PRT	1388	AA
QC	O914M4;				

```

DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
DE TYPE: XV COLLAGEN.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Helenaki P., Rehn M., Honkanen N., Myers J.C.,
RA Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding gene.";
RL J. Biol. Chem. 269:4773-4779(1994).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=98316357; PubMed=9651385;
RA Hagg P.M., Nuora A., Lilestad J., Kivirikko S., Pihlajaniemi T.;
RT "Complete exon-intron organization of the human gene for the alpha 1
RT chain of type XV collagen (COL15A1) and comparison with the homologous
RT COL18A1 gene.";
RL J. Biol. Chem. 273:17824-17831(1998).
DR EMBL: L25280; AAC78500.1; -.
DR EMBL: AF052956; AAC78500.1; JOINED.
DR EMBL: AF052957; AAC78500.1; JOINED.
DR EMBL: AF052958; AAC78500.1; JOINED.
DR EMBL: AF052959; AAC78500.1; JOINED.
DR EMBL: AF052960; AAC78500.1; JOINED.
DR EMBL: AF052961; AAC78500.1; JOINED.
DR EMBL: AF052962; AAC78500.1; JOINED.
DR EMBL: AF052963; AAC78500.1; JOINED.
DR EMBL: AF052964; AAC78500.1; JOINED.
DR EMBL: AF052965; AAC78500.1; JOINED.
DR EMBL: AF052966; AAC78500.1; JOINED.
DR EMBL: AF052967; AAC78500.1; JOINED.
DR EMBL: AF052968; AAC78500.1; JOINED.
DR EMBL: AF052969; AAC78500.1; JOINED.
DR EMBL: AF052970; AAC78500.1; JOINED.
DR EMBL: AF052971; AAC78500.1; JOINED.
DR EMBL: AF052972; AAC78500.1; JOINED.
DR EMBL: AF052973; AAC78500.1; JOINED.
DR EMBL: AF052974; AAC78500.1; JOINED.
DR EMBL: AF052975; AAC78500.1; JOINED.
DR EMBL: L25285; AAC78500.1; JOINED.
DR EMBL: L25284; AAC78500.1; JOINED.
DR EMBL: L25283; AAC78500.1; JOINED.
DR EMBL: L25282; AAC78500.1; JOINED.
DR EMBL: L25281; AAC78500.1; JOINED.
DR HSSP: P39061; IROE.
DR InterPro: IPR000087; -.
DR InterPro: IPR001791; -.
DR InterPro: IPR003129; -.
DR Pfam: PF01391; Collagen; 4.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamC; 1.
KW Collagen.
SQ SEQUENCE 1388 AA; 141756 MW; 96828E45E847194B CRC64;

Query Match 55.8%; Score 193; DB 4; Length 1388;
Best Local Similarity 49.2%; Pred. No. 5.9e-15;
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0.

1 ADRAAVPIYNLEDELFSWEALFSGSEPLKPGARIFSGDKQVLRHPTPOKSWHGS 60
|:|:|||||:|:|:|:|:| 1 |:|||:|:|:|:|:|:|
Db 1268 AERYSLPIYNLKGVLPFNWWSIFSGHGQFMNHRIPITSFDGRDITDPSNPKYIMHGS 1327

OY 61 DPN 63
OY 1:
Db 1328 SPH 1330

```

```

RESULT 10
Q9V509 PRELIMINARY: PRT: 581 AA.
ID 09V509:
AC 09V509:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG8645 PROTEIN.
GN CG8645.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Burtis K.C., Bussan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson C., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.F., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacch J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003560; AAF50621.1; -.
DR HSSP: P39061; IKOE.
DR FLYbase: Fgn0035732; CG8645.
DR InterPro: IPR000087; -.
DR InterPro: IPR002088; -.
DR Pfam: PF01391; Collagen; 3.
DR PROSITE: PS00904; PPTA; UNKNOWN.1.
SQ SEQUENCE 581 AA: 60772 MW: 19EC1E48CB477E7 CRC64;

```

```

Query Match 49.0%; Score 169.5; DB 5; Length 581;
Best local Similarity 49.2%; Pred. No. 1.7e-12;
Matches 31; Conservative 11; Mismatches 20; Indels 1; Gaps 1;
OY 1 ADRAAVPIVNLKDELLFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60

```

```

DB 429 ADR-DLPVATRGDVLFPNSKGIFFGCGFFSQAPRTISFGKNTMDTSWPKNTWHS 487
OY 61 DPN 63
DB 488 LPN 490
RESULT 11
ID 017866 PRELIMINARY: PRT: 650 AA.
AC 017866:
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE F39H11.4 OR CLE-1.
GN F39H11.4 OR CLE-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RL White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierrey Mieg J., Thomas K., Vaadin M., Vaughan K., Waterston R.,
RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.*;
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Kramer J.M., Ackley B.D.;
RT *Loss of endostatin domain from C. elegans type XVIII collagen
RT homologous causes cell and axon guidance defects.*;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z81079; CAB03084.1; -.
DR HSSP: AF164859; AAD47825.1; -.
DR InterPro: IPR000087; -.
DR Pfam: PF01391; Collagen; 2.
SQ SEQUENCE 650 AA: 69596 MW: 6CF29ED9C16B170E CRC64;

```

```

Query Match 40.5%; Score 140; DB 5; Length 650;
Best local Similarity 47.2%; Pred. No. 7.7e-09;
Matches 25; Conservative 9; Mismatches 17; Indels 2; Gaps 1;
OY 8 IVNLKDELLFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60
DB 525 VVNAVGHLPFSWRSFVNGAO--MNPALKLPSFRHVDLNDSPDKRWHS 575
RESULT 12
ID 0909K6 PRELIMINARY: PRT: 778 AA.
AC 0909K6:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CLE-1B PROTEIN.
GN CLE-1.

```

DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DR_01-JUN-1998 (TREMBLrel. 06, Last sequence update)

Search completed: September 10, 2001, 13:50:18
Job time: 108 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:48:45 ; Search time 10.07 Seconds

(without alignments)
214.309 Million cell updates/sec

Title: US-09-822-540A-1

Sequence: 346 1 ADRAVPVYNKDELLFPSPW.....DYLHPHPQKSVHWGSDPN 63

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SWISSPROT_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.6	100.0	1516	1	CAH1_HUMAN
2	29.4	85.0	1527	1	CAH1_MOUSE
3	19.3	55.8	1388	1	CAH1_HUMAN
4	63.5	18.4	266	1	CB21_SYNL
5	61	17.6	683	1	ACG1_KULU
6	61	17.6	1045	1	PRTS_SERMA
7	58.5	16.6	1720	1	FTSH_CHLVU
8	57.5	16.6	553	1	FXC1_HUMAN
9	56.5	16.5	770	1	TRPG_MOUSE
10	56.5	16.3	884	1	TP67_MYCTU
11	56.5	16.3	884	1	TP67_MYCTU
12	56.5	16.2	651	1	STM_METUA
13	55.5	16.0	727	1	NUAM_HUMAN
14	55.5	16.0	1174	1	KCRF_STRPU
15	55.5	15.9	339	1	XAP5_HUMAN
16	55.5	15.9	380	1	OP54_DROPS
17	55.5	15.9	440	1	Y433_SCHPO
18	55.5	15.9	444	1	PARA_CANFA
19	55.5	15.9	545	1	AIRE_HUMAN
20	55.5	15.9	683	1	KPCL_RAT
21	54.5	15.8	1861	1	MAP2_RAT
22	54.5	15.6	215	1	COAT_PMV
23	54.5	15.6	434	1	KP58_MOUSE
24	54.5	15.6	436	1	KP58_MOUSE
25	54.5	15.6	494	1	VATB_PLAFA
26	54.5	15.6	682	1	KPCL_HUMAN
27	54.5	15.6	983	1	YOG4_CAEL
28	53.5	15.5	705	1	Y450_RHISM
29	53.5	15.3	436	1	KP58_HUMAN
30	53.5	15.3	549	1	SVR_ARCFU
31	52.5	15.2	695	1	TKT_RHIME
32	52.5	15.2	74	1	SOM1_YEAST
33	52.5	15.2	317	1	GCH_RAT

34	52.5	15.2	350	1	FLA1_TREPA	P18193	treponema p
35	52.5	15.2	363	1	DCAM_SPIOI	P46255	spiniella ol
36	52.5	15.2	366	1	YBA8_BACCI	P48843	bacillus ci
37	52.5	15.2	372	1	OPRD_MOUSE	P32300	mus musculus
38	52.5	15.2	376	1	MTS1_RHIME	O30569	herpes simp
39	52.5	15.2	489	1	VHS_HSV11	P10225	herpes simp
40	52.5	15.2	713	1	ACG1_YEAST	O01574	saccharomyc
41	52.5	15.2	808	1	PLD_TOBAC	P93400	nicotiana t
42	52.5	15.0	267	1	CR22_ARATH	P04778	arabidopsis
43	52.5	15.0	347	1	FOS_CYPCA	P79702	cyprinus ca
44	52.5	15.0	407	1	Y116_MYCTU	P72052	mycobacteri
45	52.5	15.0	415	1	VE2_PABVE	P11329	european el

ALIGNMENTS

RESULT 1
ID CAH1_HUMAN STANDARD: PRT: 1516 AA.
AC P39060; 09Y608; 09Y607; 09UK38;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-98164096; PubMed-9503365;
RX Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
*Complete primary structure of two variant forms of human type XVIII
RT collagen and tissue-specific differences in the expression of the
RL corresponding transcripts.*;
RL Matrix Biol. 16:319-328(1998).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE-20289799; PubMed-10830953;
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Knuch J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramer J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,
RA Lehman H., Reinhardt R., Yaspo M.-L.;
The DNA sequence of human chromosome 21.;
RL Nature 405:311-319(2000).
[3]
RN SEQUENCE OF 834-1516 FROM N.A.
RP MEDLINE-94245237; PubMed-8188291;
RX Oh S.-P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
*Cloning of cDNA and genomic DNA encoding human type XVIII collagen
and localization of the alpha 1(XVII) collagen gene to mouse
RT chromosome 10 and human chromosome 21.*;
RL Genomics 19:494-499(1994).
[4]
RN SEQUENCE OF 1334-1516 FROM N.A.
RP TISSUE-Placenta;
RX Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Pu W.;
*Cloning and expression of human endostatin gene in Escherichia
RT coli.*;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
[5]
RN INVOLVEMENT IN KNOBLOCH SYNDROME.

[illegible]

FT	DISULEID	1366	1506	BY SIMILARITY.
FT	DISULEID	1466	1498	BY SIMILARITY.
FT	SITE	1095	1097	CELL ATTACHMENT SITE (POTENTIAL).
FT	VARSPLIC	1	160	MISSING (IN SHORT ISOFORM).
FT	VARSPLIC	181	215	HITTEGATLPADPPSPSGRWAPALIGSVPPSS -> MA
FT	PARSPPLIC			PCTPMPWRRRRLDVLVPLVLLGVRAASAP (IN
FT				SHORT ISOFORM).
FT	CONFLICT	428	428	F -> S (IN REF. 2).
FT	CONFLICT	841	841	I -> V (IN REF. 2).
FT	CONFLICT	877	877	V -> L (IN REF. 3).
FT	CONFLICT	886	886	P -> R (IN REF. 3).
FT	CONFLICT	912	912	P -> R (IN REF. 3).
FT	CONFLICT	933	933	R -> L (IN REF. 3).
FT	CONFLICT	975	975	P -> L (IN REF. 3).
FT	CONFLICT	1064	1064	A -> P (IN REF. 3).
FT	CONFLICT	1084	1084	L -> K (IN REF. 3).
FT	CONFLICT	1120	1120	P -> A (IN REF. 3).
FT	CONFLICT	1123	1123	P -> A (IN REF. 3).
FT	CONFLICT	1126	1126	P -> PPGP (IN REF. 2).
FT	CONFLICT	1206	1206	G -> GQ (IN REF. 3).
FT	CONFLICT	1304	1304	R -> G (IN REF. 3).
FT	CONFLICT	1314	1314	A -> G (IN REF. 3).
FT	CONFLICT	1323	1324	LR -> CG (IN REF. 3).
FT	CONFLICT	1437	1437	D -> N (IN REF. 4).
FT	CONFLICT	1443	1443	R -> T (IN REF. 4).
FT	CONFLICT	1483	1483	S -> Y (IN REF. 4).
SO	SEQUENCE	1516	AA: 153840	MM: 3C70E29A4476EB76 CRC64:

Query Match	100.0%	Score 346;	DB 1;	Length 1516;
Best Local Similarity	100.0%	Pred. No. 3,16-33;		
Matches 63; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 2
CA1H_MOUSE STANDARD: PRT; 1527 AA.
ID CA1H_MOUSE
AC P39061: Q62002: Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN=BA1B/C; TISSUE=Liver;
RX MEDLINE=94245707; PubMed=8186673;
RA Rehm M.V., Hintikka E., Pihlajaniemi T.;
RT *Primary structure of the alpha 1 chain of mouse type XVII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen
RT chain.*;
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RN SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RA Rehm M., Hintikka E., Pihlajaniemi T.;
RT *Characterization of the mouse gene for the alpha-1 chain of type
RT XVII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.*;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.

```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Helinmaki P., Rehn M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 exon-intron organization in the 3' region of the corresponding
 gene.";
 RL J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Murgaki Y., Abe N., Nihomlya Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 non-triple helical domain with a tandem repeat structure and homology
 to alpha 1(XVIII) collagen.";
 RL J. Biol. Chem. 269:4042-4046(1994).
 RN [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 1(XV), characterized by extensive interruptions in the triple-helical
 region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
 CC SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
 CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; L25286; AAA58429.1;
 CC DR EMBL; D21230; BAA04762.1;
 CC DR EMBL; L01697; -; NOT_ANNOTATED_CDS.
 CC DR MIM; 120325; -;
 CC DR InterPro; IPR000087; -;
 CC DR Pfam; PF01391; Collagen; 4;
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Cell adhesion; Collagen; Glycoprotein; Signal.
 CC SIGNAL 1 25
 CC CHAIN 26 1388 COLLAGER ALPHA 1(XV) CHAIN.
 CC FT DOMAIN 26 555 NONHELIICAL REGION 1 (NC1).
 CC FT 556 573 TRIPLE-HELICAL REGION 1 (COL1).
 CC FT DOMAIN 574 618 NONHELIICAL REGION 2 (NC2).
 CC FT 619 732 TRIPLE-HELICAL REGION 2 (COL2).
 CC FT DOMAIN 733 763 NONHELIICAL REGION 3 (NC3).
 CC FT 764 798 TRIPLE-HELICAL REGION 3 (COL3).
 CC FT DOMAIN 799 822 NONHELIICAL REGION 4 (NC4).
 CC FT 823 867 TRIPLE-HELICAL REGION 4 (COL4).
 CC FT DOMAIN 868 878 NONHELIICAL REGION 5 (NC5).
 CC FT 879 949 TRIPLE-HELICAL REGION 5 (COL5).
 CC FT DOMAIN 950 983 NONHELIICAL REGION 6 (NC6).
 CC FT 984 1013 TRIPLE-HELICAL REGION 6 (COL6).
 CC FT DOMAIN 1014 1027 NONHELIICAL REGION 7 (NC7).
 CC FT 1028 1045 TRIPLE-HELICAL REGION 7 (COL7).
 CC FT DOMAIN 1046 1052 NONHELIICAL REGION 8 (NC8).
 CC FT 1053 1107 TRIPLE-HELICAL REGION 8 (COL8).
 CC FT DOMAIN 1108 1117 NONHELIICAL REGION 9 (NC9).
 CC FT 1118 1132 TRIPLE-HELICAL REGION 9 (COL9).
 CC FT DOMAIN 1133 1388 NONHELIICAL REGION 10 (NC10).

FT DOMAIN 358 555 4 X TANDEM REPEATS.
 FT REPEAT 358 408 1.
 FT REPEAT 409 459 2.
 FT REPEAT 460 509 3.
 FT REPEAT 510 555 4.
 FT CARBOHYD 306 306 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 687 687 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 814 814 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 1046 1046 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CONFLICT 10 10 C -> S (IN REF. 2).
 FT CONFLICT 49 49 D -> V (IN REF. 2).
 FT CONFLICT 95 95 L -> A (IN REF. 2).
 FT CONFLICT 150 150 P -> A (IN REF. 2).
 FT CONFLICT 204 204 M -> V (IN REF. 2).
 FT CONFLICT 409 409 R -> A (IN REF. 2).
 SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;

Query Match 55.8%; Score 193; DB 1; Length 1388;
 Best Local Similarity 49.2%; Pred. No. 4.5e-15;
 Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNKLDELFPSEWALFSGSEGLPKRGARIFSGKDVLRHPTWPKSVHGS 60
 DB 1268 AERKSLPVLNKGQVLFNNWDSIFSGHGQFMNHIPYISFGDRDITDPSWPKVYHGS 1327
 OY 61 DPN 63
 DB 1328 SPH 1330

RESULT 4
 ID CB21_SIGNAL STANDARD; PRT; 266 AA.
 AC P13851;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CHLOROPHYLL A-B BINDING PROTEIN 1 PRECURSOR (LHCII TYPE I CAB-1)
 DE (LHCP).
 OS Sinapis alba (White mustard) (Brassica hirta).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Sinapis.
 OX NCBI_TaxID=3728;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Chloplast;
 RX MEDLINE=92322955; PubMed=1623179;
 RA Gailly A., Bartschauer A., von Arnim A., Koesel H.;
 RT "Isolation and characterization of a gene encoding a chlorophyll a/b-
 RT binding protein from mustard and the targeting of the encoded protein
 RT to the thylakoid membrane of pea chloroplasts in vitro.";
 RL Plant Mol. Biol. 19:277-287(1992).
 CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
 CC RECEPTOR. IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
 CC WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
 CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
 CC GRANA MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
 CC OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDIATE THE
 CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
 CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS A & B AND
 CC CHLOROPHYLL A-B BINDING PROTEINS.
 CC CHLOROPHYLL A-B BINDING LOCATIONS: CHLOROPLAST THYLAKOID MEMBRANE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X15894; CAA33903.1; -
 CC EMBL: X16436; CAA34459.1; -
 DR PIR: S06765; S06765.
 DR PIR: S22511; S22511.
 DR InterPro: IPR001344; -
 DR Pfam: PF00504; chloro_a-b-bind: 1.
 DR Chlorophyll: Photosynthesis: Photosystem I; Photosystem II;
 DR Thylakoid membrane: Chloroplast; Transit peptide: Multigene family;
 KW Thylakoid membrane: Chloroplast; Transit peptide: Multigene family;
 KW Transmembrane; Phosphorylation.
 FT TRANSIT 1 34 CHLOROPLAST (PROBABLE).
 FT CHAIN 35 266 CHLOROPHYL A-B BINDING PROTEIN 1.
 FT TRANSMEM 99 118 POTENTIAL.
 FT TRANSMEM 151 171 POTENTIAL.
 FT TRANSMEM 220 236 POTENTIAL.
 SQ SEQUENCE 266 AA; 28232 MW; 63BCD583D846AB7C CRC64;

Query Match 18.4%; Score 63.5; DB 1; Length 266;
 Best Local Similarity 39.0%; Pred. No. 1.9;
 Matches 16; Conservative 4; Mismatches 18; Indels 3; Gaps 1;

QY 24 FSGSEGLPKGARIFFSGKDVLR--HPTWPKSVHSGD 61
 DB 13 FACKAVKLSPGASEVFCTGRVTKTKVPTGSGSPWYGS 53

RESULT 5
 ID ACST_KLUOLA STANDARD; PRT; 683 AA.
 AC 060011;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ACETYL-COENZYME A SYNTHETASE 1 (EC 6.2.1.1) (ACETATE-COA LIGASE 1)
 DE (ACYL-ACTIVATING ENZYME 1).
 GN ACST.
 OS Kluyveromyces fragilis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
 RA Zeeman A.M., Henkel C., Steensma H.Y.;
 RU Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE REQUIRED FOR ASSIMILATION OF ETHANOL AND ACETATE.
 CC -1- CATALYTIC ACTIVITY: ATP + ACETATE + COA -> AMP + PYROPHOSPHATE +
 CC ACETYL-COA.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF061265; AAC16713.1; -
 CC InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding: 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00455; AMP_BINDING: 1.
 KW Lysase.
 SQ SEQUENCE 683 AA; 76029 MW; 7AC8876A497966DB CRC64;

Query Match 17.6%; Score 61; DB 1; Length 683;
 Best Local Similarity 27.1%; Pred. No. 11;

Matches 16; Conservative 8; Mismatches 13; Indels 22; Gaps 3;

QY 25 SGSEGLPKGARIFFSGKDVLRHPT
 DB 460 AGGATPMKPGAAFPFGDLAVDPTTGTEGCHAEGLAIKRWPSFARTIMNND 518

RESULT 6
 ID PRTS_SERMA STANDARD; PRT; 1045 AA.
 AC P09489;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE EXTRACELLULAR SERINE PROTEASE PRECURSOR (EC 3.4.21.-).
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-35 AND 407-408.
 RC STRAIN=IFO 3046;
 RX MEDLINE=86223815; PubMed=3011754;
 RA Yangida N., Uozumi T., Beppu T.;
 RT *Specific excretion of Serratia marcescens protease through the outer
 RT membrane of Escherichia coli.*
 RT J. Bacteriol. 166:937-944(1986).
 RN [2]
 RP PARTIAL SEQUENCE, AND PROCESSING.
 RX MEDLINE=92348352; PubMed=1639760;
 RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;
 RT *Detection of large COOH-terminal domains processed from the
 RT precursor of Serratia marcescens serine protease in the outer
 RT membrane of Escherichia coli.*
 RT J. Biochem. 111:627-632(1992).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M13469; AAA26572.1; -
 DR PIR: A29840; A29840.
 DR HSP; Q09405; IMPT.
 DR MEROPS; S08.094; -
 DR InterPro: IPR000209; -
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.
 FT PROPEP 646 1045
 FT ACT_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 1045 AA; 112345 MW; 4924EA50B4FF179C CRC64;

Query Match 17.6%; Score 61; DB 1; Length 1045;
 Best Local Similarity 34.9%; Pred. No. 17;
 Matches 15; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 14 ELFPSPMEALFSGSEGLPKGARIFFSGKDVLRHPTWPKSV 56

DB 449 ECFSDSWNDISGHGILTKTGAGTLALLGNNTYRGDTWVAGGV 491

RESULT 7

FTSH_CHLVU STANDARD; PRT; 1720 AA.
 AC P56369;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CELL DIVISION PROTEIN FTSH HOMOLOG.
 GN FTSH.
 OS Chlorella vulgaris.
 OC Chlorophyta.
 OC Eukaryota: Viridiplantae: Chlorophyta: Trebouxiophyceae: Chlorellales:
 OC Chlorellaceae: Chlorella.
 NC NCBITaxid=3077;
 RN (1)

RC SEQUENCE FROM N.A.
 RC STRAIN-TAM C-27 / TAMIVA.
 RA MEDLINE-97303241; PubMed-9159184;
 RA Tsubaki T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
 RA Tsubaki T., Nakashima K., Tsubaki T., Suzuki Y., Hamada A., Ohta T.,
 RA Tsubaki T., Tsubaki K., Sugita M.,
 RA "Complete nucleotide sequence of the chloroplast genome from the
 RT green alga Chlorella vulgaris: the existence of genes possibly
 RT involved in chloroplast division."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).

-1- SUBCELLULAR LOCATION: CHLOROPLAST; INTEGRAL MEMBRANE PROTEIN
 (POTENTIAL).
 -1- DOMAIN: LACKS THE ZINC PROTEASE DOMAIN OF OTHER FTSH PROTEINS.
 -1- ALSO MUCH LONGER IN BOTH THE N- AND C-TERMINUS.
 -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: AB001684; BAA57905.1; -;
 DR EMBL: AB001684; BAA57905.1; ALT_INIT.
 DR InterPro: IPR001939; -;
 DR Pfam: PF000004; AAA; 1.
 DR PROSITE: PS00674; AAA; 1.
 KW Cell division; ATP-binding; Transmembrane; Chloroplast.
 FT TRANSMEM 48 68
 FT TRANSMEM 896 916 POTENTIAL.
 FT TRANSMEM 973 993 POTENTIAL.
 FT TRANSMEM 1021 1041 POTENTIAL.
 SQ SEQUENCE 1720 AA; 197172 MW; 0F1RA926B799D5BB CRC64;

Query Match 16.9%; Score 58.5; DB 1; Length 1720;

Best Local Similarity 26.7%; Pred. No. 57;
 Matches 16; Conservative 14; Mismatches 21; Indels 9; Gaps 4;

QY 9 VNLKDELFF-PSWEALFSGSEGLPKP-----GARIFSGDKVLRHPTW-PQKSVHGS 60
 DB 1480 IGFEQIRYSSHWQEDVS-AEMERKPNPKGSLIYDERTSRNENYPPDEFTHNS 1538

RESULT 8

FXCI_HUMAN STANDARD; PRT; 553 AA.
 AC Q12948; GUP06;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-
 DE RELATED TRANSCRIPTION FACTOR 3) (FREAC-3).

GN FOXCI OR FKHL7 OR FREAC3.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 NC NCBITaxid=9606;
 RN (1)

RP SEQUENCE FROM N.A., AND VARIANTS S-112; M-126 AND L-131.
 RA MEDLINE-98282091; PubMed-9620769;
 RA Nishimura D.Y., Swiderski R.E., Alward W.L.M., Searby C.C.,
 RA Pail S.R., Bennet S.R., Kanis A.B., Gastier J.M., Stone E.M.,
 RA Sheffield V.C.;
 RT "The forkhead transcription factor gene FKHL7 is responsible for
 RT glaucoma phenotypes which map to 6p25."
 RL Nat. Genet. 19:140-147(1998).

RN (2)
 RP SEQUENCE FROM N.A., AND VARIANTS ARA THR-82 AND MET-87.
 RX MEDLINE-99011252; PubMed-9792859;
 RA Meers A.J., Jordan T., Mirzayans F., Dubois S., Kume T., Parlee M.,
 RA Ritch R., Koop B., Kuo W.-L., Collins C., Marshall J., Gould D.B.,
 RA Pearce W., Carlsson P., Enerbaeck S., Morissette J., Bhattacharya S.,
 RA Hogen B., Raymond V., Walter M.A.;
 RT "Mutations of the forkhead/winged-helix gene, FKHL7, in patients with
 RT Axenfeld-Rieger anomaly."
 RL Am. J. Hum. Genet. 63:1316-1328(1998).

RN (3)
 RP SEQUENCE OF 73-178 FROM N.A.
 RX MEDLINE-95045392; PubMed-7957066;
 RA Pierron S., Helqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;
 RT "Cloning and characterization of seven human forkhead proteins:
 RT binding site specificity and DNA bending."
 RL EMBO J. 13:5002-5012(1994).

-1- FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES
 RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.

CC -1- SUBUNIT: MONOMER.
 CC -1- DISEASE: DEFECTS IN FKHL7 ARE THE CAUSE A SPECTRUM OF GLAUCOMA
 CC PHENOTYPES SUCH AS AXENFELD-RIEGER ANOMALY (ARA), AXENFELD-RIEGER
 CC SYNDROME (ARS) AND IRIDOGONIOSGENESIS ANOMALY (IGDA). ARS IS AN
 CC AUTOSOMAL DOMINANT DISORDER PRESENTING WITH ARA-LIKE OCULAR
 CC FINDINGS IN ADDITION TO ABNORMALITIES OF THE TEETH, JAW AND
 CC UMBILICUS.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: AF048693; AAC18081.1; -;
 DR EMBL: U13221; AAA92038.1; -;
 DR EMBL: AF078096; AAC72915.1; -;
 DR MIM: 601090; -;
 DR MIM: 601631; -;
 DR InterPro: IPR001766; -;
 DR InterPro: IPR002952; -;
 DR InterPro: IPR002965; -;
 DR Pfam: PF00250; Fork_head; 1.

DR PRINTS: PR00053; FORKHEAD.
 DR PRINTS: PR01217; PRICHEXTNSN.
 DR PRINTS: PR01228; EGGSHL.
 DR PROSITE: PS00657; FORK_HEAD_1; 1.
 DR PROSITE: PS00658; FORK_HEAD_2; 1.
 DR PROSITE: PS50039; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation;
 KW Disease mutation.
 FT DNA_BIND 77 168 FORK-HEAD.
 FT DOMAIN 28 33 POLY-ALA.
 FT DOMAIN 169 173 POLY-ARG.
 FT DOMAIN 194 197 POLY-PRO.
 FT DOMAIN 262 272 POLY-SER.

```

FT DOMAIN 292 297 POLY-PRO.
FT DOMAIN 375 382 POLY-GLY.
FT DOMAIN 438 443 POLY-SER.
FT DOMAIN 447 456 POLY-GLY.
FT DOMAIN 486 495 POLY-ALA.
FT VARIANT 82 82 S -> T (IN AAA).
FT VARIANT 87 87 /FTID-VAR_007944.
FT VARIANT 112 112 I -> M (IN AAA).
FT VARIANT 126 126 /FTID-VAR_007945.
FT VARIANT 131 131 I -> S (IN IGDA).
FT VARIANT 131 131 /FTID-VAR_007815.
FT VARIANT 131 131 I -> M (IN AAA).
FT VARIANT 131 131 /FTID-VAR_007816.
FT VARIANT 131 131 S -> L (IN ARS).
FT VARIANT 131 131 /FTID-VAR_007817.
FT VARIANT 131 131 V -> L (IN REF. 2).
FT VARIANT 131 131 ROPP -> ASPR (IN REF. 2).
FT VARIANT 131 131 N -> D (IN REF. 2).
SO SEQUENCE 553 AA; 56787 MW; DBC81B943030359E CRC64;

```

```

Query Match 16.6%; Score 57.5; DB 1; Length 553;
Best Local Similarity 37.9%; Pred. No. 22;
Matches 22; Conservative 2; Mismatches 29; Indels 5; Gaps 2;

```

```

QY 4 AAVPIVNLKDELFPSEALFSGSEGP-LKPGARISFDCKDVLHPTPOKSWHGS 60
DB 252 AAVPIKESPPS---SSSSLSGSSPPGSLPSARPLSLDGDADNAPPAPAPPPHHS 305

```

```

RESULT 9
KPC L MOUSE STANDARD; PRT; 683 AA.
AC P23258;

```

```

DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROTEIN KINASE C, ETA TYPE (EC 2.7.1.1) (NPKC-ETA) (PKC-L).
GN PRKCH OR PKCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;

```

```

RN RP SEQUENCE FROM N.A.
RC TISSUE-Epidermis;
RX MEDLINE-91093089; PubMed-2266135;
RA Osada S.I., Mizuno K., Salido T.C., Akita Y., Suzuki K., Kuroki T.,
RA Ohno S.;

```

```

RT *A phorbol ester receptor/protein kinase, nPKC eta, a new member of
RT the protein kinase C family predominantly expressed in lung and
RT skin.
RT J. Biol. Chem. 265:22434-22440(1990).

```

```

CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LUNG AND SKIN.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

DR EMBL: D90242; BAA14288.1; -.
DR PIR: A23690; A23690.
DR HSSP: P28867; 1PTR.
DR MGD: MGI:97600; PKCH.
DR InterPro: IPR000008; -.
DR InterPro: IPR000719; -.
DR InterPro: IPR000961; -.
DR InterPro: IPR002219; -.
DR InterPro: IPR002290; -.
DR Pfam: PF00168; C2_1.
DR Pfam: PF00130; DAG-pe_bind; 2.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00433; pkinase; 1.
DR PRINTS: PR00008; DAGEDOMAIN.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS50049; DAG-pe_bind_DOM_1; 2.
DR PROSITE: PS50081; DAG-pe_bind_DOM_2; 2.
DR PROSITE: PS50107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50108; PROTEIN_KINASE_SF; 1.
DR KAP-binding; Transferase; Serine/threonine-protein kinase;
KW Phorbol-ester binding; Zinc; Duplication.
FT DOMAIN 12 102 C2 DOMAIN.
FT DOMAIN 172 222 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 246 295 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 355 614 PROTEIN KINASE.
FT NE_BIND 361 369 ATP (BY SIMILARITY).
FT BINDING 384 384 ATP (BY SIMILARITY).
FT ACT_SITE 479 479 BY SIMILARITY.
SO SEQUENCE 683 AA; 77973 MW; C7DB338A9F95F576 CRC64;

```

```

Query Match 16.5%; Score 57; DB 1; Length 683;
Best Local Similarity 31.6%; Pred. No. 32;
Matches 18; Conservative 8; Mismatches 15; Indels 16; Gaps 4;

```

```

QY 11 LKDELFPSEALFSGSEGPL-----PGARISF---DGKVLHPTPOKSWHGS 53
DB 567 LKDELFPSEALFSGSEGPL-----PGARISF---DGKVLHPTPOKSWHGS 53

```

```

RESULT 10
TRPG ASPNG STANDARD; PRT; 770 AA.
ID TRPG ASPNG
AC P05328;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT II (EC 4.1.3.27) [INCLUDES: GLUTAMINE
DE AMIDOTRANSFERASE; INDOL-3-GLYCEROL PHOSPHATE SYNTHASE (EC 4.1.1.48)
DE (IGPS); N-(5'-PHOSPHORIBOSYL)ANTHRANILATE ISOMERASE (EC 5.3.1.24)
DE (PRAI)].
DE TRPG.

```

```

OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OC NCBI_Taxid=5061;
RN (1)
RN RP SEQUENCE FROM N.A.
RC STRAIN=401;
RX MEDLINE-88223483; PubMed-2836085;
RX Kos T., Kuijvenhoven A., Heessing H.G.M., Pouwels P.H.,
RX van den Hondel C.A.M.J.J.;

```

```

RT *Nucleotide sequence of the Aspergillus niger trpC gene: structural
RT relationship with analogous genes of other organisms.
RT Curr. Genet. 13:137-144(1988).

```

```

[2]
SEQUENCE OF 1-69 AND 392-433 FROM N.A.
RX MEDLINE-86137391; PubMed-2936650;
RX Kos A., Kuijvenhoven J., Wernars K., Bos C.J., van den Broek H.W.J.,
RX Pouwels P.H., van den Hondel C.A.M.J.J.;
RT *Isolation and characterization of the Aspergillus niger trpC gene.
RT Gene 39:231-238(1985).

```



```

CC -1- FUNCTION: TRIFUNCTIONAL ENZYME BEARING THE GLN AMIDOTRANSFERASE
CC (GATASE) DOMAIN OF ANTHRANILATE SYNTHASE, INDOL-GLYCEROLPHOSPHATE
CC SYNTHASE, AND PHOSPHORIBOSYLANTHRANILATE ISOMERASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: N-5'-PHOSPHORIBOSYL-ANTHRANILATE =
CC 1-(2-CARBOXY-5-AMINO-1-DEOXY-D-RIBULOSE 5-PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: 1-(2-CARBOXY-5-AMINO-1-DEOXY-D-RIBULOSE
CC 5-PHOSPHATE + 1-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE + CO(2) + H(2O).
CC -1- PATHWAY: FIRST, THIRD AND FOURTH STEPS OF TRYPTOPHAN BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X07071; CAA30107.1; -
CC DR EMBL: M14403; AAA32710.1; -
CC DR EMBL: M14404; AAA32709.1; -
CC DR PIR: S00643; S00643.
CC DR PIR: A23979; A23979.
CC DR HSSP: 006121; 1JUK.
CC DR INTERPRO: IPR000991; -
CC DR INTERPRO: IPR001240; -
CC DR INTERPRO: IPR001468; -
CC DR INTERPRO: IPR002385; -
CC DR Pfam: PF00117; GATase; 1.
CC DR Pfam: PF00118; IGPS; 1.
CC DR Pfam: PF00697; PRAI; 1.
CC DR PRINTS: PR00096; GATASE.
CC DR PRINTS: PR00097; ANTSNTHASE1.
CC DR PROSITE: PS00442; GATASE_TYPE_1; 1.
CC DR PROSITE: PS00614; IGPS; 1.
CC KM tryptophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
CC Decarboxylase; Transferase; Glutamine amidotransferase.
CC FT DOMAIN 25 219 GLUTAMINE AMIDOTRANSFERASE.
CC FT DOMAIN 255 519 INDOL-3-GLYCEROL PHOSPHATE SYNTHASE.
CC FT DOMAIN 535 770 N-(5'-PHOSPHORIBOSYL)ANTHRANILATE
CC ISOMERASE.
CC FT ACT_SITE 104 104 GATASE (BY SIMILARITY).
CC FT ACT_SITE 199 199 GATASE (BY SIMILARITY).
CC FT ACT_SITE 201 201 GATASE (BY SIMILARITY).
CC SQ SEQUENCE 770 AA: 82909 MM: 36DBDE5B23097012 CRC64;

Query Match 16.3%; Score 56.5; DB 1; Length 770;
Best Local Similarity 30.6%; Pred. No. 41;
Matches 15; Conservative 7; Mismatches 26; Indels 1; Gaps 1;

OY 2 DRANAIVMLKDELLFPSSWDALEFSGSEGLKPGARIFSGDKVLRHPT 50
DB 572 DVVALPIQSVVRSTPKPASQALHTSOEPRAATSVVEYFDHSK-ILMHP 619

RESULT 11
YF67_MYCTU STANDARD; PRT; 884 AA.
AC Q50654; Q50731;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 95.4 KDA PROTEIN RV2567.
OS RV2567 OR MYCY227.34C OR MYCY9CA.01C.
GN Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetiales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Sulton J.E.,
RA Taylor K., Whitehead S., Barrall B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS RV2411C AND SYNECHOCYSTIS PCC
CC 6803 SL10335.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: Z77250; CAB01053.1; -
CC DR Tuberculat; RV2567; -
CC KM Hypothetical protein.
CC SQ SEQUENCE 884 AA: 95448 MM: 95D23A4DZEDDB365 CRC64;

Query Match 16.3%; Score 56.5; DB 1; Length 884;
Best Local Similarity 30.2%; Pred. No. 48;
Matches 16; Conservative 11; Mismatches 13; Indels 13; Gaps 2;

OY 23 LFGSGEGLKPGARIFSGD-----KDVLRHPT-----PKSVHMSDP 62
DB 620 LMSLVDPDRPSLQSVSEGLAQAQVARDQLSDPTMYLVANVERAVERHSDP 672

RESULT 12
SYM_METUA STANDARD; PRT; 651 AA.
AC Q50659;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE-TRNA LIGASE)
DE (METERS).
GN METG OR M01263.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Welton J.F., Funtman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Huzar M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
CC TRNA(FMET) AMINOACYLATION.
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +

```



```
CC      PYRROPHOSPHATE + L-METHIONYL-TRNA(MET).
CC      -I- COPROCTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC      -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC      -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC      STRONG; TO CYSTEINYL-TRNA SYNTHETASE.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to licensel@isb-sib.ch).
CC      -----
CC      EMBL: U67567; AAB99269.1; -.
CC      DR      HSSP: P00959; IMED.
CC      DR      TIGR: M1263; -.
CC      DR      InterPro: IPR001412; -.
CC      DR      InterPro: IPR002300; -.
CC      DR      InterPro: IPR002304; -.
CC      DR      InterPro: IPR002547; -.
CC      DR      Pfam: PF00133; tRNA-synt_1; 1.
CC      DR      Pfam: PF01588; tRNA_bind_1.
CC      DR      PRINTS: PR01041; TRNASYNTHMET.
CC      *DR      PROSITE: PS00178; AA-TRNA_LIGASE_1; FALSE_NEG.
CC      KW      aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
CC      tRNA-binding; metal-binding; zinc.
CC      FT      SITE           10       20          "HIGH" REGION.
CC      FT      SITE           320     324          "KMSKS" REGION.
CC      FT      DOMAIN         556     649          tRNA BINDING.
CC      FT      METAL          142     142          ZINC (BY SIMILARITY).
CC      FT      METAL          145     145          ZINC (BY SIMILARITY).
CC      FT      METAL          154     154          ZINC (BY SIMILARITY).
CC      FT      METAL          157     157          ZINC (BY SIMILARITY).
CC      SO      SEQUENCE      651 AA; 75558 MW; C5C58DF5A5A591E07 CRC64;

Query Match              16.2%; Score 56; DB 1; Length 651;
Best Local Similarity    37.5%; Pred. No. 40;
Matches 12; Conservative 5; Mismatches 13; Indels 2; Gaps 1;

QY      31 LKPGARIFFSDGKDVLRRPT--WPKSVVHGS 60
        I: : : : | | | : | : | : | : | : | : |
-DL      270 LEKDKIVHFICGDITVHHAVFWPMGLAHGS 301

RESULT 13
NTAM_HUMAN
.AC      p28331;
.DT      01-DEC-1992 (rel. 24, Created)
.DT      01-DEC-1992 (rel. 24, Last sequence update)
.DT      15-JUL-1999 (rel. 38, last annotation update)
.DE      NADH+UBIQUINONE OXIDOREDUCTASE 75 KDA SUBUNIT PRECURSOR (EC 1.6.5.3)
.DE      (EC 1.6.99.3) (COMPLEX I-75SKD) (CI-75SKD).
.GN      NDUFSL1.
.OS      Homo sapiens (Human).
.OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
.OX      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
.OX      NCBI_Taxid=9606;
.RN      [1]
.RP      SEQUENCE FROM N.A.
.RX      MEDLINE=92037608; PubMed=1935949;
.RA      Chow W., Ragan J., Robinson B.H.;
.SB      "Determinatlon of the cDNA sequence for the human mitochondrial
RT      75-kDa Fe-S protein of NADH-coenzyme Q reductase."
RL      Eur. J. Biochem. 201:547-550(1991).
CC      -I- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A
CC      COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT
CC      MAY FORM PART OF THE ACTIVE SITE CREVICE WHERE NADH IS OXIDIZED.
CC      -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC      -I- CORRECTOR: MAY BIND ONE 2FE-2S CLUSTER & ONE 4FE-4S CLUSTER.
CC      -I- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
```

```
CC -1- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
CC MITOCHONDRION INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KDA SUBUNIT FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR DR EMBL: X61100; CAA43412.1; -.
DR PIR: S16382; S16382.
DR PIR: S17854; S17854.
DR MIM: 157655; -.
DR InterPro: IPR000283; -.
DR InterPro: IPR001041; -.
DR InterPro: IPR001467; -.
DR Pfam: PF00111; ferz_1.
DR Pfam: PF00364; molybdopterin_1.
DR PROSITE: PS00641; COMPLEXI_75K_1; 1.
DR PROSITE: PS00642; COMPLEXI_75K_2; 1.
DR PROSITE: PS00643; COMPLEXI_75K_3; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transist peptide;
KW Iron-sulfur; 4Fe-4S.
FT TRANSIT 1 23
FT CHAIN 24 727
FT METAL 53 53
FT METAL 54 53
FT METAL 63 64
FT METAL 64 64
FT METAL 75 75
FT METAL 78 78
FT METAL 92 92
FT METAL 128 128
FT METAL 131 131
FT METAL 137 137
FT METAL 176 176
FT METAL 179 179
FT METAL 182 182
FT METAL 226 226
SQ SEQUENCE 727 AA; 79573 MW; 1C175EC156BFSF CRC64;
Query Match 16.0%; Score 55.5; DB 1; Length 727;
Best Local Similarity 40.9%; Pred. No. 51;
Matches 18; Conservative 3; Mismatches 16; Indels 7; Gaps 2;
OY 3 RAAPIV-NLKDELLPFWSEALFGSGSEPLKPGAIRISFDCKDV 45
DB 301 RLTPWVRNEKGILLTYTSMEDALSRYAVGLD-----SFQGDV 338
| : | | | | | | | | | | | | | | | | | | | |
RESULT 14
ID _KCRF_STRPU STANDARD; PRT: 1174 AA.
AC P18296;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CREATINE KINASE, FLAGELLAR (BC 2.7.3.2).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidae; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Testis;
RX MEDLINE=90311370; PubMed=2367531;
RA Woche D.O.; Chardonneau H.; Shapiro B.M.;
RT "The phosphocreatine shuttle of sea urchin sperm: flagellar creatine
kinase resulted from a gene triplication.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 87:5203-5207(1990).
CC -1- FUNCTION: THIS AXONEMAL PROTEIN PARTICIPATES IN AN ENERGY SHUTTLE
CC THAT UTILIZES PHOSPHOCREATINE TO TRANSFER THE ENERGY FROM ATP
CC GENERATED BY THE MITOCHONDRION IN THE SPERM HEAD TO DYNEIN IN THE
CC DISTAL PORTIONS OF THE FLAGELLUM.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES SPECIFICALLY WITH THE AXONEME AND
CC MAY BIND DIRECTLY TO POLYMERIZED MICROTUBULES.
CC -1- DOMAIN: CONTAINS THREE COMPLETE BUT NONIDENTICAL CREATIVE KINASE
CC SEGMENTS FLANKED BY UNIQUE REGIONS.
CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M33763; AAA30049.1; -.
CC PIR: A43736; A43736.
CC DR HSPD; P11009; ICRK.
CC DR InterPro: IPR000749; -.
CC DR Pfam: PF00217; ATP-gua_ptnans; 3.
CC DR PROSITE: PS00112; GUANIDO_KINASE; 3.
CC DR KW Transferrase; Kinase; Repeat.
CC FT REPEAT 61 414 1 (APPROXIMATE).
CC FT REPEAT 434 787 2 (APPROXIMATE).
CC FT REPEAT 808 1161 3 (APPROXIMATE).
CC FT ACT_SITE 324 324 BY SIMILARITY.
CC FT ACT_SITE 697 697 BY SIMILARITY.
CC FT ACT_SITE 1071 1071 BY SIMILARITY.
CC FT ACT_SITE 1174 AA; 130869 MW; B8F4FBA5AC48EE93 CRC64;
CC SO SEQUENCE

```

```

Query Match 16.0%; Score 55.5; DB 1; Length 1174;
Best Local Similarity 31.0%; Pred. No. 86;
Matches 13; Conservative 4; Mismatches 14; Indels 11; Gaps 2;

```

```

OY 23 LEFSGEGLPKGARIFSDGKQVLRHPTWPO-KSYWHSQDPN 63
DB 607 LEFKPVSPPLTAARM-----HROMPGRGIMHNENKN 638

*RESULT 15
XAP5_HUMAN STANDARD; PRT; 339 AA.
ID XAP5_HUMAN
AC 014320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE XAP-5 PROTEIN (HXC-26 PROTEIN).
GN XAP5 OR HXC26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RP MEDLINE=97480734; PubMed=9339379;
RA Mazzarella R., Pengue G., Yoon J., Jones J., Schlessinger D.;
RT "Differential expression of XAP5, a candidate disease gene.";
RL Genomics 45:216-219(1997).
RN 12
RP SEQUENCE OF 15-339 FROM N.A.
RP TISSUE=Sketetal muscle;
RX PubMed=9039504;
RA Toyoda A., Sakai T., Sugiyama Y., Kusuda J., Hashimoto K., Maeda H.;
RT "Isolation and analysis of a novel gene, HXC-26, adjacent to the rab
RT GDP dissociation inhibitor gene located at human chromosome Xq28
RT region.";
RL DNA Res. 3:337-340(1996).

```

```

RN 13
RP SEQUENCE OF 66-339 FROM N.A.
RX PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccocioppa A., Chen C.-N.,
RA Zuo L., Helner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: fifteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RN Hum. Mol. Genet. 5:659-668(1996).
CC -1- FUNCTION: MAY BE A DNA-BINDING PROTEIN OR TRANSCRIPTIONAL FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED. MOSTLY
CC ABUNDANT IN FETAL BRAIN, LIVER AND KIDNEY; IN THE ADULT, HIGH
CC LEVELS WERE ALSO OBSERVED IN HEART, SKELETAL MUSCLE, SPLEEN,
CC THYMUS, PROSTATE AND SMALL INTESTINE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AD001530; AAB81663.1; -.
CC DR EMBL: D83389; BA11907.1; -.
CC DR EMBL: D83261; BA11907.1; JOINED.
CC DR EMBL: D83388; BA11907.1; JOINED.
CC DR EMBL: D83260; BA11871.1; -.
CC DR EMBL: L4140; AAA92649.1; -.
CC KW Nuclear protein.
CC FT DOMAIN 152 155 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SO SEQUENCE 339 AA; 40241 MW; 88BCA57EA9B0AA32 CRC64;

```

```

Query Match 15.9%; Score 55; DB 1; Length 339;
Best Local Similarity 23.6%; Pred. No. 26;
Matches 21; Conservative 11; Mismatches 15; Indels 42; Gaps 5;

```

```

OY 8 IYNLDELFPSEVALFS-----GSEGPLKPGARIFSD----- 41
DB 249 IMYIKEDLLIPHHHSFYDFIVTKARGKSGPL-----FNFVDVDDVRLLSDATVERDESH 302
OY 42 -GKQVLRHPTWPOK-----SYWHSQDP 62
DB 303 AKGVLR--SWYEKNKHIFPASHMEYDP 329

```

Search completed: September 10, 2001, 13:50:35
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:48:05 ; Search time 13.46 Seconds
(without alignments)
356.538 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346
Sequence: 1 ADRAAVPIVIMKDELFPSPW.....DVLHPTPQKSWHGSDPN 63

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	684	2 A53019	collagen alpha 1(X
2	294	85.0	1774	2 B56101	collagen alpha 1(X
3	291	84.1	1315	2 A56101	collagen alpha 1(X
4	193	55.8	1388	2 A53317	collagen alpha 1(X
5	140	40.5	650	2 T22002	hypothetical prote
6	67.5	19.5	614	2 S27962	modulator recognit
7	64	18.5	644	2 A72919	probable 2-oxoacid
8	63.5	18.4	266	2 S22511	chlorophyll a/b-bi
9	63	18.2	336	2 T17408	rRNA (adenine-N6-)
10	62.5	18.1	467	2 T26705	hypothetical prote
11	62.5	18.1	620	2 T69797	conserved hypotet
12	62	17.9	419	2 T25565	hypothetical prote
13	61	17.6	1045	2 A29840	serine proteinase
14	60	17.3	984	2 T44496	cellulose 1,4-beta
15	59.5	17.2	268	2 H75614	hypothetical prote
16	59.5	17.2	382	2 S24434	class I histocompa
17	59	17.1	303	2 T41056	conserved hypotet
18	58.5	16.9	305	2 S07115	class I histocompa
19	58.5	16.9	354	2 T80170	class I histocompa
20	58.5	16.9	481	2 T23131	hypothetical prote
21	58.5	16.9	1184	2 D86387	probable protein p
22	58.5	16.9	1720	2 T07258	cell division prot
23	58	16.8	461	2 A54024	protein kinase (EC
24	58	16.8	662	2 T41215	probable acetate-
25	58	16.8	2055	2 T30259	multiple PDZ domai
26	57.5	16.6	506	2 H64618	sigma-54 interacti
27	57	16.5	683	1 A23690	protein kinase (EC
28	57	16.5	1228	1 JC5573	copper-transporlin
29	57	16.5	2054	2 T46612	multi PDZ domain p

30	56.5	16.3	211	2 T31724	hypothetical prote
31	56.5	16.3	217	2 T27331	hypothetical prote
32	56.5	16.3	770	2 S00643	anthranilate synth
33	56.5	16.3	884	2 C70729	hypothetical prote
34	56	16.2	651	2 F64457	methionine--C-RNA 1
35	56	16.2	753	1 D72660	probable aldehyde
36	56	16.2	838	2 T45699	hypothetical prote
37	55.5	16.0	299	2 D75138	hypothetical prote
38	55.5	16.0	553	2 T35451	ABC transporter in
39	55.5	16.0	574	2 T48113	inositol-1,4,5-tri
40	55.5	16.0	727	1 S17854	NADH dehydrogenase
41	55.5	16.0	1174	1 A43736	creatine kinase (E
42	55.5	16.0	1568	2 T08616	aggregation factor
43	55	15.9	274	2 S24439	class I histocompa
44	55	15.9	322	2 H86405	probable oxidoredu
45	55	15.9	325	2 JC5276	HXC-26 protein - h

ALIGNMENTS

RESULT 1

A53019

collagen alpha 1(XVIII) chain - human (fragment)

N:Contains: endostatin

C:Species: Homo sapiens (man)

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #ext_change 31-Mar-2000

C:Accession: A53019

R:Oh, S.P.; Worman, M.L.; Seidlin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse

Genomics 19, 494-499, 1994

A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local

A:Reference number: A53019; MVID:94245237

A:Accession: A53019

A:Molecule type: mRNA

A:Residues: 1-684 <COHA>

A:Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794

A:Note: The cited accession number, U25548, is not in Genbank release 103

A:Note: In the authors' translation, 482-Gly is not shown, residues 483-490 are shift

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

lated and subsequently O-glycosylated.

C:Comment: different splice forms of collagen alpha 1(XVIII) may be involved in periv

C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of

ay be useful in treating solid tumors.

C:Genetics:

A:Gene: GDB:COL18A1

A:Cross-references: GDB:138752; OMIM:120328

A:Map position: 21q22.3-21q22.3

C:Superfamily: unassigned collagens

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog

F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>

F:1-59/Domain: collagenous (fragment) #status predicted <CO5>

F:74-115/Domain: collagenous #status predicted <CO5>

F:129-201/Domain: collagenous #status predicted <CO6>

F:212-244/Domain: collagenous #status predicted <CO7>

F:257-278/Domain: collagenous #status predicted <CO8>

F:262-264/Region: cell attachment (R-G-D) motif

F:286-340/Domain: collagenous #status predicted <CO9>

F:354-371/Domain: collagenous #status predicted <CO10>

F:502-684/Product: endostatin #status predicted <EST>

F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 100.0%; Score 346; DB 2; Length 684;

Best local similarity 100.0%; Pred. No. 86-34;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVIMKDELFPSPWALFSGSEGPLKPRARIFSDGKVLRLHPTPQKSWHGS 60

DB 565 ADRAAVPIVIMKDELFPSPWALFSGSEGPLKPRARIFSDGKVLRLHPTPQKSWHGS 624

QY 61 DPN 63

DB 625 DPN 627

RESULT 2

B56101

Collagen alpha 1(XVIII) chain precursor, long splice form - mouse
 M:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
 C:Accession: B56101; C56101; S72450; S65595; P06675; A54072; A58816

R:Rehm, M.; Pihlajaniemi, T.
 J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and classification homologous to rat and Drosophila frizzled proteins.
 A:Reference number: A56101; MUID:95181468

A:Accession: B56101

A:Molecule type: mRNA

A:Residues: 1-562 <REH1>

A:Cross-references: GB:U11637; NID:9618429; PIDN:AAC52179.1; PID:9618430
 A:Experimental source: splice form clone PE17.24

A:Accession: C56101

A:Molecule type: mRNA

A:Residues: 1-239, 487-562 <REH2>

A:Cross-references: GB:U11637; NID:9618429

A:Experimental source: splice form clones PE8.1, PE19, PE15.2

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 submitted to the EMBL Data Library, August 1993

A:Reference number: S72450

A:Accession: S72450

A:Molecule type: mRNA

A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16

A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa

A:Reference number: A58370; MUID:94240111

A:Accession: S65595

A:Molecule type: mRNA

A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>

A:Cross-references: EMBL:L22545

R:Abel, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Nishimura, Y.
 Biochem. Biophys. Res. Commun. 196, 576-582, 1993

A:Title: Identification of a novel collagen chain represented by extensive interruptions

A:Reference number: P06675; MUID:94059075

A:Accession: P06675

A:Molecule type: mRNA

A:Residues: 635-1774 <AB2>

R:Rehm, M.; Hinkley, E.; Pihlajaniemi, T.
 J. Biol. Chem. 269, 13929-13935, 1994

A:Title: Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial str

collagen chain.

A:Reference number: A54072; MUID:94245707

A:Accession: A54072

A:Molecule type: DNA

A:Residues: 1293-1403, 'R', 1405-1774 <REH3>

A:Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734

R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bl

Cell 88, 277-285, 1997

A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.

A:Reference number: A58816; MUID:97160848

A:Accession: A58816

A:Molecule type: protein

A:Residues: 1591-1610 <ORE>

A:Experimental source: hemangioendothelium cells

A:Note: Inhibits endothelial cell proliferation

A:Comment: Prolins and lysines at the third position of the tripeptide repeating unit

lated and subsequently O-glycosylated

C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per

C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un

ay be useful in treating solid tumors.

C:Genetics:

A:Gene: MGI:COL18a1

A:Cross-references: MGI:71175

A:Map position: 10:41.0

A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15

A:Note: the list of introns is incomplete

C:Superfamily: unassigned collagens

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog

F:1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status p

F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form

F:1-24/Domain: signal sequence #status predicted <SIG>

F:361-486/Region: frizzled similarity

F:786-812/Domain: collagenous #status predicted <CO01>

F:823-896/Domain: collagenous #status predicted <CO02>

F:921-1042/Domain: collagenous #status predicted <CO03>

F:1066-1148/Domain: collagenous #status predicted <CO04>

F:1163-1204/Domain: collagenous #status predicted <CO05>

F:1218-1290/Domain: collagenous #status predicted <CO06>

F:1301-1333/Domain: collagenous #status predicted <CO07>

F:1346-1369/Domain: collagenous #status predicted <CO08>

F:1351-1353/Region: cell attachment (R-G-D) motif

F:1377-1428/Domain: collagenous #status predicted <CO09>

F:1442-1459/Domain: collagenous #status predicted <CO10>

F:1591-1774/Product: endostatin #status predicted <EST>

F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity

F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 85.0%; Score 294; DB 2; Length 1774;

Best Local Similarity 81.0%; Pred. No. 4,7e-27;

Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAVPYINLDELLEFSEALFSGSECPKPGARIFSFQDKVLRHPYPOKSWHGS 60

DB 1654 ADRCGVPIYNLKDVELVSLPSRWSLFSGSGQLOPGARIFSFQDKVLRHPYPOKSWHGS 1713

QY 61 DPN 63

DB 1714 DPS 1716

RESULT 3 4

A56101

Collagen alpha 1(XVIII) chain precursor, short splice form - mouse

M:Contains: endostatin

C:Species: Mus musculus (house mouse)

C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000

C:Accession: A56101; A58371; S72450; S65595

R:Rehm, M.; Pihlajaniemi, T.

J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and cl

if homologous to rat and Drosophila frizzled proteins.

A:Reference number: A56101; MUID:95181468

A:Accession: A56101

A:Molecule type: mRNA

A:Residues: 1-103 <REH1>

A:Cross-references: GB:U11636; NID:9618427; PIDN:AAC52178.1; PID:9618428

R:Rehm, M.; Pihlajaniemi, T.

Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno

A:Reference number: A58371; MUID:94240112

A:Accession: A58371

A:Molecule type: mRNA

A:Residues: 1-928 <REH2>

A:Cross-references: GB:U11698; NID:9404754; PIDN:AA37434.1; PID:9553894

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

submitted to the EMBL Data Library, August 1993

A:Reference number: S72450

A:Accession: S72450

A:Molecule type: mRNA

A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHW>

A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-

A:Reference number: A58370; MUID:9424011
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 28-1315 <OHS>
A:Cross-references: EMBL:L22545
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are acetylated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in perlecanin synthesis.
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of unglycosylated endostatin.
C:Genetics:
G:Gene: MGI:COL18A1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglycan
F:1-25/Domain: signal sequence #status predicted <SIG>
F:24-235/Region: thrombospondin amino-terminal similarity
F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M>
F:327-355/Domain: collagenous #status predicted <CO1>
F:364-437/Domain: collagenous #status predicted <CO2>
F:462-583/Domain: collagenous #status predicted <CO3>
F:607-689/Domain: collagenous #status predicted <CO4>
F:704-745/Domain: collagenous #status predicted <CO5>
F:759-831/Domain: collagenous #status predicted <CO6>
F:842-874/Domain: collagenous #status predicted <CO7>
F:887-910/Domain: collagenous #status predicted <CO8>
F:892-894/Region: cell attachment (R-G-D) motif
F:918-969/Domain: collagenous #status predicted <CO9>
F:983-1000/Domain: collagenous #status predicted <CO10>
F:1137-1315/Product: endostatin #status predicted <EST>
F:1138-1315/Region: multiplexin collagen carboxy1-terminal similarity
E:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:172-228/Disulfide bonds: #status predicted
F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 84.1%; Score 291; DB 2; Length 1315;
Best Local Similarity 79.4%; Pred. No. 7,7e-27;
Matches 50; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

OY 1 ADNAAPVIVNLKDELFPSSWALFSGSSEGLPKPGRIRISFDCKVDLRHPTQSKSWHGSGS 60
||| :||||||| | |||:|||||:| :||||||| ||||| |||||
DB 1195 ADNGSVPIVNLKDVLSPSMDSLFSSGGGVQGPARIFSDGRDYLRLHPAMPQKSVMHGSGS 1254

OY 61 DPN 63
|||
DB 1255 DPS 1257

RESULT 4
A53317
N:Alternative names: procollagen alpha 1(XV) chain
C:Species: Homo sapiens (man)
C>Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Mar-2000
A:Accession: A53317; A53146; S28778
R:Kiivirikko, S.; Heilmann, F.; Rehn, M.; Honkanen, N.; Myers, J.C.; Phlajantemi, T.
J. Biol. Chem. 269, 4773-4779, 1994
A>Title: Primary structure of the alpha chain of human type XV collagen and exon-intron organization.
A:Reference number: A53317; MUID:94148920
A:Accession: A53317
A:Molecule type: preliminary
A>Status: preliminary
A:Residues: 1-1388 <KITV>
A:Cross-references: GB:L25280
A>Note: nucleotide sequence and conceptual translation not complete
J. Muragaki, Y.; Abe, N.; Nimomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A>Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple helix domain.
A:Reference number: A53146; MUID:94140817
A:Accession: A53146

A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-9,'S','11'-48,'V','50-94,'A','96-149,'A','151-203,'V','205-408,'A','410-569 '<M>
C:CROSS-references: GB:D21230; MID:9415605; PIDN:BAA04762.1; PID:d1005294; RID:g46070
R:Werners,J.C.; Kivirikko,S.; Gordon,M.R.; Dion,A.S.: Philajanteml,T.
Proc.Natl.Acad.Sci.U.S.A. 89, 10144-10148, 1992
A>Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha
A:Reference number: S28778; MUID:93066196
A:Accession: S28778
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 544-640,'P','642-811,'P','813-1252 <MYE>
C:Genetics:
A:Gene: GDB:COL15A1
A:CROSS-references: GDB:132578; OMIM:120325
A:Map position: 9q21-q9q22
C:Superfamily: unassigned collagens
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 55.8%; Score 193; DB 2; Length 1388;
Best Local Similarity 49.2%; Pred.No. 5.7e+15;
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;

OY 1 ADRAAPVTNLKDELLFPSEALFSGSEGPKLGARIFSPDCKDVLRHPMPQKSWMHGS 60
 :|::||| |:: || ::||| | I |::|||:: ||: |::|||
Db 1268 AERYSLPTVLNKGVLFNWDSIFSGHGGOFGFMHIPIYSDFGRIDMTDPSPQKVWHGS 1327

OY 61 DPN 63
 :|:
Db 1328 SPH 1330

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22002
R:white, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19500
A:Accession: T22002
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-650 <MIT>
A:CROSS-references: EMBL:Z81079; PIDD:CA803084.1; GSPPDB:GN00019; CESP:F39H11.4
A:Experimental source: clone F39H11
C:Genetics:
A:Gene: CESP:F39H11.4
A:Map position: 1
A:introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 40.5%; Score 140; DB 2; Length 650;
Best Local Similarity 47.2%; Pred.No. 5.9e-09;
Matches 25; Conservative 9; Mismatches 17; Indels 2; Gaps 1;

OY 8 INLKDELFPSEALFSGSEGPKLGARISSFDGKDYLRRHTPMPOKSWMHGS 60
 :||: ||| |: |: ||:|||| | | | | | | | | | | | | | | | |
Db 525 VNVAGHHLPWSRSFYNGAQ--MNPHAKLFFSDRHVDLVNDSRWDKRVMHGS 575

RESULT 6
S27962
modulator recognition factor 1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
C:Accession: S27962
R:Okada, T.; Merrillis, B.W.; Huang, E.; Okada, T.; Tanaka, Y.; Gertson, P.; Itakura, K.
Submitted to the EMBL Data library, March 1991

RESULT 15
H75614
hypotheical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75614
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75614
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12307.1; PID:g646060
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0182
A:Map position: 2

Query Match 17.28; Score 59.5; DB 2; Length 268;
Best Local Similarity 36.1%; Pred. No. 11;
Matches 13; Conservative 7; Mismatches 15; Indels 1; Gaps 1;
OY . 13 DELFPSEALFSGSGCP-KPGARIFSPDGKDVLR 47
| : : : | : : | | | | : | | |
DB 165 DDASTAALDRLAAGCPISNPAGRIFFPENNDALR 200

Search completed: September 10, 2001, 13:49:50
Job time: 105 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:47:45 : Search time 12.18 Seconds
(without alignments)
106.502 Million cell updates/sec

Title: US-09-822-540A-1

Sequence: 1 ADRAAVPIVNLKDELLFPPSW.....DYLRLHPTWPKSVWHGSDPN 63

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTOTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	183	4	US-09-206-059-2
2	291	84.1	195	1	US-08-159-784-2
3	277.5	80.2	185	3	US-08-985-526-36
4	190	54.9	191	1	US-08-159-784-3
5	94	27.2	16	4	US-09-385-442-32
6	59	17.1	14	4	US-09-385-442-31
7	57.5	16.6	506	2	US-08-849-480A-5
8	57.5	16.6	553	3	US-09-083-351-2
9	57.5	16.6	553	4	US-09-083-352-2
10	55	15.9	444	1	US-08-483-140-28
11	55	15.9	444	2	US-08-485-938A-32
12	54.5	15.8	470	2	US-08-377-440A-1
13	54.5	15.8	578	1	US-08-653-740-3
14	54.5	15.8	578	2	US-09-073-594-3
15	54.5	15.8	578	3	US-09-275-925-3
16	54.5	15.8	636	1	US-08-653-740-5
17	54.5	15.8	636	2	US-09-073-594-5
18	54.5	15.8	636	3	US-09-275-925-5
19	53	15.3	312	4	US-09-216-295-21
20	53	15.3	371	4	US-09-104-308-1
21	52.5	15.2	293	1	US-08-628-291-4
22	52.5	15.2	293	2	US-09-128-722-4
23	52.5	15.2	317	1	US-08-628-291-12
24	52.5	15.2	317	2	US-09-128-722-12
25	52.5	15.2	372	1	US-08-149-093A-6
26	52.5	15.2	372	2	US-08-911-245-6
27	52.5	15.2	372	2	US-08-514-451A-9

28	52.5	15.2	372	2	US-08-411-859-2	Sequence 2, Appl 1
29	52.5	15.2	372	2	US-08-411-859-10	Sequence 10, Appl 1
30	52.5	15.2	372	3	US-09-170-331-6	Sequence 6, Appl 1
31	52.5	15.2	372	3	US-08-147-592A-4	Sequence 4, Appl 1
32	52.5	15.2	372	4	US-08-430-266A-6	Sequence 6, Appl 1
33	52	15.0	449	3	US-08-987-743-7	Sequence 7, Appl 1
34	52	15.0	514	4	US-08-796-899-25	Sequence 25, Appl 1
35	52	15.0	841	4	US-09-413-814-107	Sequence 107, Appl 1
36	51	14.7	444	2	US-08-485-938A-33	Sequence 33, Appl 1
37	50.5	14.6	164	2	US-08-918-723-3	Sequence 3, Appl 1
38	50.5	14.6	164	2	US-09-237-507-3	Sequence 3, Appl 1
39	50.5	14.6	294	1	US-08-628-291-2	Sequence 2, Appl 1
40	50.5	14.6	294	1	US-09-128-722-2	Sequence 2, Appl 1
41	50.5	14.6	318	1	US-08-628-291-16	Sequence 16, Appl 1
42	50.5	14.6	318	2	US-09-128-722-16	Sequence 16, Appl 1
43	50.5	14.6	1235	2	US-08-680-326-36	Sequence 36, Appl 1
44	50	14.5	31	3	US-09-045-632-77	Sequence 77, Appl 1
45	50	14.5	98	3	US-09-045-632-5	Sequence 5, Appl 1

ALIGNMENTS

RESULT 1
US-09-206-059-2
; Sequence 2, Application US/09206059

Patent No. 6201104
GENERAL INFORMATION:

APPLICANT: Macdonald, Nicholas
APPLICANT: Sim, Kim Lee
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
FILE REFERENCE: 05213-0370
CURRENT APPLICATION NUMBER: US/09/206,059
CURRENT FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-09-206-059-2

Query Match 100.0%; Score 346; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 2e-38;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPPSWALFSGSEGPLKPGARIFSDGKDVLRHPTWPKSVWHGS 60
DB 64 ADRAAVPIVNLKDELLFPPSWALFSGSEGPLKPGARIFSDGKDVLRHPTWPKSVWHGS 123

QY 61 DPN 63
DB 124 DPN 126

RESULT 2
US-08-159-784-2
; Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 84.1% Score 291; DB 1; Length 195;
Best Local Similarity 79.4%; Pred. No. 4,1e-31;
Matches 50; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 60
DB 75 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 134
QY 61 DPN 63
DB 135 DPS 137

RESULT 3
US-08-985-526-36
Sequence 36, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Boyle, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorow Jr., Robert G
TELECOMMUNICATION INFORMATION:

TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 80.2% Score 277.5; DB 3; Length 185;
Best Local Similarity 76.6%; Pred. No. 2.4e-29;
Matches 49; Conservative 10; Mismatches 4; Indels 1; Gaps 1;
QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 59
DB 64 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 123
QY 60 SDPN 63
DB 124 SDPS 127

RESULT 4
US-08-159-784-3
Sequence 3, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 191
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-3

Query Match 54.9% Score 190; DB 1; Length 191;
Best Local Similarity 49.2%; Pred. No. 9.9e-18;
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;
QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 60
DB 75 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSWHGS 134

APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
ZIP: 02109-2170
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,351
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-7000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-351-2

Query Match 16.6%; Score 57.5; DB 3; Length 553;
Best Local Similarity 37.9%; Pred. No. 14;
Matches 22; Conservative 2; Mismatches 29; Indels 5; Gaps 2;

QY 4 AAVPIVNLKDELLFSPWEALFSGSGEP-LKPGARIFSPGKDYLRHPTPQKSWHGS 60
DB 252 AAVPKIESPDS-----SSSSLSGSGSPGSLPSARPLSLDGADSAAPPAPPPHHS 305

RESULT 9
US-09-083-352-2
Sequence 2, Application US/09083352
Patent No. 6207450
GENERAL INFORMATION:
APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: GLUCOCOMA THERAPEUTICS AND DIAGNOSTICS
TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
ZIP: 02109-2170
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,352
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-7000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-352-2

Query Match 16.6%; Score 57.5; DB 4; Length 553;
Best Local Similarity 37.9%; Pred. No. 14;
Matches 22; Conservative 2; Mismatches 29; Indels 5; Gaps 2;

QY 4 AAVPIVNLKDELLFSPWEALFSGSGEP-LKPGARIFSPGKDYLRHPTPQKSWHGS 60
DB 252 AAVPKIESPDS-----SSSSLSGSGSPGSLPSARPLSLDGADSAAPPAPPPHHS 305

RESULT 10
US-08-483-140-28
Sequence 28, Application US/08483140
Patent No. 5698403
GENERAL INFORMATION:
APPLICANT: ICOS Corporation
TITLE OF INVENTION: Platelet-Activating Factor Acetyl
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gertein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,140
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318,905
FILING DATE: 6-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,803
FILING DATE: 6-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 5698403and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32781
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3658
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-140-28

Query Match 15.9%; Score 55; DB 1; Length 444;
Best Local Similarity 30.8%; Pred. No. 23;
Matches 12; Conservative 7; Mismatches 20; Indels 0; Gaps 0;

OY 17 FPSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKS 55
DB 403 PDQWDLIEGKDNLMPTGNTNITNEHDTLONSPEAKS 441

RESULT 11
US-08-485-938A-32
Sequence 32, Application US/08485938A
Patent No. 5847088
GENERAL INFORMATION:
APPLICANT: Cousens, Lawrence S.
APPLICANT: Eberhardt, Christine D.
APPLICANT: Gray, Patrick W.
APPLICANT: Le Trong, Hai
APPLICANT: Tjoelker, Larry W.
TITLE OF INVENTION: Platelet-Activating Factor
TITLE OF INVENTION: Acetylhydrolase
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485.938A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318.905
FILING DATE: 06-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133.803
FILING DATE: 06-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 5847088and, Greta E.
REGISTRATION NUMBER: 35.302
REFERENCE/DOCKET NUMBER: 27866/32792
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3658
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-938A-32

Query Match 15.9%; Score 55; DB 2; Length 444;
Best Local Similarity 30.8%; Pred. No. 23;
Matches 12; Conservative 7; Mismatches 20; Indels 0; Gaps 0;
OY 17 FPSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKS 55

DB 403 PDQWDLIEGKDNLMPTGNTNITNEHDTLONSPEAKS 441

RESULT 12
US-08-377-440A-1
Sequence 1, Application US/08377440A
Patent No. 5985623
GENERAL INFORMATION:
APPLICANT: POLLOCK, THOMAS J.
APPLICANT: YAMAZAKI, MOTOHIDE
APPLICANT: THORNE, LINDA
APPLICANT: MIKOLAJCZAK, MARCIA
APPLICANT: ARMENTROUT, RICHARD W.
TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: JULES E. GOLDBERG
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/377.440A
FILING DATE: 24-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24.408
REFERENCE/DOCKET NUMBER: JG-SEC-1166
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-377-440A-1

Query Match 15.8%; Score 54.5; DB 2; Length 470;
Best Local Similarity 36.8%; Pred. No. 28;
Matches 14; Conservative 8; Mismatches 7; Indels 9; Gaps 2;

OY 9 VNIKDELFPSWEALFSGSEGPLKPGARIFSPDGKDVLR 46
DB 234 NNIKEELIVQPNAL--GAIG-----VDSYEGKDTL 262

RESULT 13
US-08-653-740-3
Sequence 3, Application US/08653740
Patent No. 5792850
GENERAL INFORMATION:
APPLICANT: James W. Baumgartner
APPLICANT: Donald C. Foster
APPLICANT: Frank J. Grant
APPLICANT: Cindy A. Sprecher
TITLE OF INVENTION: HEMATOPOIETIC CYTOKINE RECEPTOR
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East

Db	85	AGRSWVAIPREQLTMSDKLLVGTGKAGQPLMPVFVNLETOMKPNAPRLGPDVDFSEDDP	144
OY	46	LR-----HPTWPKSV	56
Db	145	LEATVHWAPPTWPSHKV	161

Search completed: September 10, 2001, 13:49:29
Job time: 104 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:47:25 ; Search time 20.16 Seconds

(without alignments)
189,450 Million cell updates/sec

Title: US-09-822-540A-1

Sequence: 1 ADRAAVPIVNLKDELFPSW.....DVLHPTWPKSVWHGSDPN 63

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Database : A_Geneseq_0601.*

1:	/SID88/gcgdata/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	178	21	AAV94324
2	346	100.0	182	21	AAV94324
3	346	100.0	182	21	AAV94323
4	346	100.0	182	21	AAV94323
5	346	100.0	183	20	AAV94322
6	346	100.0	183	20	AAV94322
7	346	100.0	183	20	AAV94322
8	346	100.0	183	21	AAV94321
9	346	100.0	183	21	AAV94321
10	346	100.0	183	21	AAV94321
11	346	100.0	183	22	AAV94321

12	346	100.0	193	21	AAV90877	Human HMW endostat
13	346	100.0	195	21	AAV90874	Human HMW endostat
14	346	100.0	216	21	AAV90495	Amino acid sequenc
15	346	100.0	684	18	AAV26327	Human alpha-1 coll
16	346	100.0	684	20	AAV25113	Human alpha-1 (XVI)
17	346	100.0	1301	20	AAV92296	Human collagen 18
18	346	100.0	1336	20	AAV08694	Human collagen 18
19	342	98.8	183	22	AAV9810	Human endostatin p
20	335	96.8	271	21	AAV08407	A human anglogen
21	302	87.3	184	21	AAV70265	Canine anglogen
22	284	85.0	108	22	AAV9607	Murine endostatin
23	284	85.0	184	22	AAV9380	Murine endostatin
24	294	85.0	191	21	AAV28398	Murine endostatin
25	293	84.7	184	20	AAV18409	Endostatin protein
26	291	84.1	184	20	AAV08689	Murine endostatin
27	291	84.1	184	21	AAV70258	Murine anglogen
28	291	84.1	207	22	AAV71830	Murine endostatin
29	291	84.1	218	20	AAV08691	Murine gene therap
30	291	84.1	580	20	AAV08692	Murine gene therap
31	291	84.1	684	20	AAV25114	Mouse alpha-1 (XVI)
32	291	84.1	1288	18	AAV26328	Mouse alpha-1 coll
33	291	84.1	1288	20	AAV92297	Mouse alpha-1 (XVI)
34	277.5	80.2	185	20	AAV06197	Anti-angiogenic en
35	255	73.7	184	22	AAV9381	Chicken endostatin
36	255	56.9	180	22	AAV9383	Murine endostatin
37	193	55.8	176	22	AAV90875	Human HMW endostat
38	193	55.8	180	22	AAV9382	Human endostatin S
39	193	55.8	181	20	AAV22227	Human reslin prote
40	193	55.8	181	20	AAV18410	Reslin protein seq
41	193	55.8	191	21	AAV90876	Human HMW endostat
42	182	52.6	45	22	AAV35586	Antiangiogenic pen
43	114	32.9	85	20	AAV22226	Apoptogen. Homo s
44	109	31.5	50	22	AAV35585	Antiangiogenic pen
45	107	30.9	20	22	AAV94956	Human endostatin p

ALIGNMENTS

RESULT 1	AAV94324	standard; Protein; 178 AA.
ID	AAV94324	
XX	AAV94324	
AC	AAV94324	
XX	AAV94324	
DT	11-AUG-2000	(first entry)
XX	11-AUG-2000	
DE	Alternate human endostatin protein.	
KW	Human: endothelial cell proliferation inhibitor; collagen XVII;	
KW	angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;	
KW	vasotropic; dermatological; ophthalmological; vulvular;	
KW	antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;	
KW	ocular angiogenic disease; atherosclerosis; scleroderma;	
KW	myocardial angiogenesis; telangiectasia; angiofibroma;	
KW	wound granulation.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200026368-A2.	
PD	11-MAY-2000.	
XX		
PF	01-NOV-1999;	99WO-US25605.
XX		
PR	30-OCT-1998;	98US-0106343.
XX		
PR	20-MAY-1999;	99US-0315689.
XX		
PA	(CHIL-) CHILRENS MEDICAL CENT.	
XX		
PI	O'Reilly MS, Folkman MJ;	
XX		
DR	WPI; 2000-365617/31.	

DR N-PSDB: AAA27005.

XX Novel endostatin capable of inhibiting endothelial cell proliferation
 PT and angiogenesis, useful for treating angiogenesis-dependent cancers
 PT and as birth control agents

PS Claim 3; Page 39; 68pp; English.

XX The present sequence is an alternate functional endostatin
 CC protein. When the human endostatin gene sequence AAA27004 is
 CC recombinantly expressed, an observable doublet of protein results, both
 CC versions of which are functional endostatin proteins. The present
 CC endostatin variant is the same as the protein encoded by AAA27004 minus
 CC the first four amino acids. Recombinant mouse endostatin (20 mg/kg) was
 CC administered subcutaneously to mice implanted with Lewis lung carcinomas.
 CC There was tumour mass regression non-detectable levels after 12 days of
 CC therapy due to the angiogenesis inhibitory activity of endostatin. Thus
 CC the protein is useful for treatment of angiogenesis-dependent cancers.
 CC The polynucleotide and polypeptide sequences of this endostatin are
 CC useful for treating and diagnosis of tumours, ocular angiogenic
 CC diseases, Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiodioma
 CC and wound granulation, for treatment of diseases related to excessive or
 CC abnormal stimulation of endothelial cells e.g. intestinal adhesions,
 CC atherosclerosis, scleroderma. The protein may also be useful as a birth
 CC control agent by reducing or preventing uterine vascularisation. The
 CC gene for endostatin may be isolated from cells or tissue that express
 CC high levels of endostatin, eg. tumour cells, by generating cDNA from
 CC mRNA using reverse transcriptase and then amplifying the DNA sequence.
 XX Sequence 178 AA:

Query Match 100.0%; Score 346; DB 21; Length 178;
 Best Local Similarity 100.0%; Pred. No. 5.5e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIYNLDELFPSEALFSGSEGPLKPGARIFSFQDKDYLRRHPTWPKSWHGS 60
 DB 60 adraavpiynlkdellfpsealfsgsegpikpgariffsqdkdylrrhptwqkswhgs 119

OY 61 DPN 63
 DB 120 dpn 122

RESULT 2

AAB28399 standard; Protein: 182 AA.

AAC AAB28399;

DT 19-FEB-2001 (first entry)

DE Human endostatin.

KM Human: endostatin; cytosolic; antiproliferative;

KM vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;

KM cancer; vascularised solid tumour.

OS Homo sapiens.

PN WO200064946-A2.

PD 02-NOV-2000.

PF 28-APR-2000; 2000WO-US11367.

PR 28-APR-1999; 99US-0131432.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Thorpe PE, Brekken RA;

XX WPI: 2000-687317/67.

XX Immunogenic composition for the treatment and diagnosis of cancer
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -

PS Example 10; Page 291-292; 298pp; English.

XX The present invention relates to anti-Vascular Endothelial Growth Factor
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
 CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
 CC receptor VEGFR1. The present sequence is human endostatin.
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.
 CC The anti-VEGF antibodies of the present invention are useful for the
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.
 XX Sequence 182 AA:

Query Match 100.0%; Score 346; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIYNLDELFPSEALFSGSEGPLKPGARIFSFQDKDYLRRHPTWPKSWHGS 60
 DB 64 adraavpiynlkdellfpsealfsgsegpikpgariffsqdkdylrrhptwqkswhgs 123

OY 61 DPN 63
 DB 124 dpn 126

RESULT 3

AAV94323 standard; Protein: 182 AA.

AAC AAV94323;

DT 11-AUG-2000 (first entry)

DE Human endostatin protein.

KM Human: endothelial cell proliferation inhibitor; collagen XVIIII;

KM angiogenesis inhibitor; anti-tumour; cytosolic; antiproliferative;

KM vasotropic; dermatological; ophthalmological; vulvatory;

KM antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;

KM ocular angiogenic disease; atherosclerosis; scleroderma;

KM myocardial angiogenesis; telangiectasia; angiodioma;

KM wound granulation.

OS Homo sapiens.

PN WO200026368-A2.

PD 11-MAY-2000.

PF 01-NOV-1999; 99WO-US25605.

PR 30-OCT-1998; 98US-0106343.

PR 20-MAY-1999; 99US-0315689.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI O'Reilly MS, Folkman MJ;

WPI: 2000-365617/31.

N-PSDB: AAA27004.

PT Novel endostatin capable of inhibiting endothelial cell proliferation
 PT and angiogenesis, useful for treating angiogenesis-dependent cancers
 PT and as birth control agents

XX Claim 2: Page 38; 68pp; English.
 PS
 CC The present sequence is an endostatin protein which is the carboxy
 CC terminal protein of human collagen XVIII. Recombinant mouse endostatin
 CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
 CC lung carcinomas. There was tumour mass regression non-detectable levels
 CC after 12 days of therapy due to the angiogenesis inhibitory activity of
 CC endostatin. Thus the protein is useful for treatment of angiogenesis-
 CC dependent cancers. The polynucleotide and polypeptide sequences of this
 CC endostatin are useful for treating and diagnosis of tumours, ocular
 CC angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiodioma and wound granulation, for treatment of diseases related to
 CC excessive or abnormal stimulation of endothelial cells e.g. intestinal
 CC adhesions, atherosclerosis, scleroderma. The protein may also be useful
 CC as a birth control agent by reducing or preventing uterine
 CC vascularisation. The gene for endostatin may be isolated from cells or
 CC tissue that express high levels of endostatin, eg. tumour cells, by
 CC generating cDNA from mRNA using reverse transcriptase and then amplifying
 CC the DNA sequence.
 CC
 SQ Sequence 182 AA:
 Query Match 100.0%; Score 346; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ADRAAPYIVNLKDELLFPSEALFSGSEGLPKPGARIFSGDKVLRHPTWPKSVWHS 60
 DB 64 adraapivnlkdeellfpsealifsgseglpkpgarifsgdkvrlrhpwtpkswhs 123
 OY 61 DPN 63
 DB 124 dpn 126
 Db 124 dpn 126
 RESULT 4
 AAY59622
 ID AAY59622 standard; protein: 182 AA.
 AC AAY59622;
 DT 14-MAR-2000 (first entry)
 DE Human endostatin protein fragment.
 KW Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;
 KW metastatic cancer; tumorigenesis; ocular angiogenic disease;
 KW rheumatoid arthritis; Osler-Webber syndrome; telangiectasia;
 KW haemophilic joint; angiodioma; wound granulation.
 OS Homo sapiens.
 PN WO9962944-A2.
 PD 09-DEC-1999.
 PF 03-JUN-1999; 99WO-US12278.
 PR 03-JUN-1998; 98US-0087890.
 PR 10-JUL-1998; 98US-0092393.
 PR 01-SEP-1998; 98US-0098790.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PI Javaherian K, Folkman MJ;
 WPI: 2000-072833/06.
 New endostatin oligomers, used for treating e.g. tumours -

PS Disclosure: Page 6; 44pp; English.
 CC
 CC This sequence is a fragment of the human endostatin protein. Endostatin
 CC is an approximately 20kD C-terminal globular domain of the collagen-like
 CC protein collagen XVIII. Protein oligomers consisting of more than one
 CC endostatin monomer have anti-tubulogenic effects and induce
 CC reorganization of the actin cytoskeleton. The oligomer has scatter factor
 CC activity. The oligomers induce the destruction of tubular lumens and
 CC elongation of cells, and inhibit tubulogenesis and tumorigenesis. The
 CC oligomers can also be used to treat metastatic cancers, tumours,
 CC rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Webber
 CC syndrome, plaque neovascularisation, telangiectasia, haemophilic
 CC joints, angiodioma and wound granulation. The oligomers can also be
 CC used to treat diseases that have angiogenesis as a pathological
 CC consequence e.g. ulcers. The endostatin oligomers can also be used to
 CC develop affinity columns for isolating antibodies or receptors. Passive
 CC antibody therapy using antibodies that specifically bind endostatin
 CC oligomers can be used to modulate morphogenic processes such as
 CC metastatic cancer as well as angiogenesis-dependent processes such as
 CC reproduction, development, wound healing, tissue repair, and
 CC angiogenesis-dependent diseases. Also, antisera directed to the Fab
 CC regions of endostatin oligomer antibodies can be administered to block
 CC the ability of endogenous endostatin oligomer antisera to bind endostatin
 CC oligomers.
 SQ Sequence 182 AA:
 Query Match 100.0%; Score 346; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ADRAAPYIVNLKDELLFPSEALFSGSEGLPKPGARIFSGDKVLRHPTWPKSVWHS 60
 DB 64 adraapivnlkdeellfpsealifsgseglpkpgarifsgdkvrlrhpwtpkswhs 123
 OY 61 DPN 63
 DB 124 dpn 126
 Db 124 dpn 126
 RESULT 5
 AAY08693
 ID AAY08693 standard; protein: 183 AA.
 AC AAY08693;
 DT 10-AUG-1999 (first entry)
 DE Human endostatin protein fragment.
 KW Plasmogen; human; angiotatin; endostatin; gene therapy; vector;
 KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
 KW tumour growth; solid tumour; diabetic retinopathy; retina.
 OS Homo sapiens.
 PN WO9926480-A1.
 PD 03-JUN-1999.
 PF 20-NOV-1998; 98WO-US24950.
 PR 20-NOV-1997; 97US-0975424.
 PA (GENE-) GENETIX PHARM INC.
 PI (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Bachelot T, Leboulch P, Pawliuk RJ;
 WPI: 1999-357696/30.
 DR N-PSDB; AAY77719.
 XX

PT Anti-angiogenic gene therapy vectors
 XX
 PS Disclosure: Page 74-75; 83pp; English.
 CC This invention describes a novel viral gene therapy vector comprising a
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
 CC from human or murine angiostatin, human or murine endostatin and
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
 CC sufficiently attenuated for use in human gene therapy. The products of
 CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
 CC ophthalmological activity. The vector is used in gene therapy for
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
 CC expresses an anti-angiogenic polypeptide. An additional use comprises
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
 CC inhibits angiogenesis in the vicinity of the retina. The vector is
 CC administered to cells *ex vivo* and then administered to the patient.
 CC
 SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 20; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELLFPEWEALFSGSEGLPKRGARIFSPDGKDYLRHPTPQKSYMHGS 60
 DB 64 adraavpivnlkdellfpwealfsgseglpkrgarifsfqdkvrlrhtpqpksvwhgs 123

OY 61 DPN 63
 DB 124 dpn 126

RESULT 6

ID AAY02113 standard; Protein: 183 AA.

XX AAY02113;

DT 16-JUL-1999 (first entry)

DE SEQ ID 76 of W09916889.

XX Angiostatin; endostatin; Interferon; thrombospondin;
 KW Interferon-inducible protein; platelet factor 4; anti-angiogenic;
 KW anti-tumor; multifunctional protein; angiogenesis-mediated disease;
 KW cancer; diabetic retinopathy; macular degeneration; arthritis;
 KW tumor cell production.

XX Homo sapiens.

XX W09916889-A1.

XX 08-APR-1999.

XX 30-SEP-1998; 98WO-US20464.

XX 01-OCT-1997; 97US-0060609.

XX (SEAR) SEARLE & CO G D.

XX Bojanowski MA, Caparon MH, Casperson GF, Gregory SA;
 PI Klein BK, McKearn JP;

XX WPI: 1999-255098/21.

PT New multifunctional proteins useful for treating angiogenic-mediated
 PT diseases

XX Disclosure: Page 106-107; 121pp; English.

CC The specification describes multifunctional proteins which comprise
 CC combinations of angiostatin, endostatin, interferon, thrombospondin,

CC Interferon-inducible protein and platelet factor 4, and have
 CC anti-angiogenic and/or anti-tumor activity. The multifunctional protein
 CC may exhibit useful properties such as having similar or greater
 CC biological activity when compared to a single factor or by having
 CC improved half-life or decreased adverse side effects, or a combination
 CC of these properties. The proteins can be used for treating an
 CC angiogenic-mediated disease, e.g. cancer, diabetic retinopathy, macular
 CC degeneration, or arthritis. They can also be used for inhibiting the
 CC production of tumor cells (characteristic of lung, breast, ovarian,
 CC prostate, pancreatic, gastric, colon, renal, bladder cancers; melanoma,
 CC hepatoma, sarcoma and lymphoma) in a patient and for inhibiting tumor
 CC growth. The present sequence is used in the course of the invention.
 CC
 SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 20; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELLFPEWEALFSGSEGLPKRGARIFSPDGKDYLRHPTPQKSYMHGS 60
 DB 64 adraavpivnlkdellfpwealfsgseglpkrgarifsfqdkvrlrhtpqpksvwhgs 123

OY 61 DPN 63
 DB 124 dpn 126

RESULT 7

ID AAB30493 standard; Protein: 183 AA.

XX AAB30493;

DT 06-MAR-2001 (first entry)

DE Amino acid sequence of human endostatin encoded by plasmid pMALCH#15.

XX Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amylase;

XX endostatin; cancer; tumour growth; angiogenesis.

XX Homo sapiens.

XX W0200060945-A1.

XX 19-OCT-2000.

XX 12-APR-2000; 2000WO-US09747.

XX 13-APR-1999; 99US-0129084.

XX (MERI) MERCK & CO INC.

XX Desanti CL, Strohl WR;

XX WPI: 2000-686970/67.

XX N-PSDB; AAC62023.

PT Preparation of soluble recombinant endostatin involves transforming
 PT Streptomyces host with expression vector comprising nucleotide
 PT sequence encoding endostatin operably linked to linker and leader
 PT peptide

XX Example 1: Fig 6; 57pp; English.

CC The present sequence represents human endostatin. The protein is
 CC expressed in Streptomyces. Leader sequences of Streptomyces sp. strain
 CC C5 SnPA and S. venezuelae alpha-amylase proteins are linked to the
 CC N-terminal of endostatin. This ensures that endostatin protein is
 CC produced as a secreted, soluble protein which needs no refolding, is
 CC stable in the fermentation broth and is produced in large quantities.
 CC The method is used for preparing soluble recombinant human, murine or

CC primate endostatin, which is useful in the treatment of cancer,
 CC inhibition of tumour growth, inhibition of angiogenesis, isolation of
 CC receptors for endostatin and for identification of anti-angiogenic
 CC compounds in assays. The endostatin protein is produced as a secreted,
 CC soluble protein which needs no refolding, is stable in the fermentation
 CC broth and is produced in large quantities. Streptomycetes are amenable
 CC for cultivation in large fermentations allowing for large quantities of
 CC soluble endostatin to be produced.
 XX
 SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5,7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIYNLKDLELFPSEALFSGSEGLPKGARIFSFDDKDVLRHPTWQKSYWHS 60
 DB 64 adraavpivnlkdellfpsealifsgseglpkgarifsfddgkdvlrhptwqkswbgs 123
 OY 61 DPN 63
 DB 124 dpn 126

RESULT 8

AAB16451
 ID AAB16451 standard; Protein; 183 AA.

AC AAB16451;

DT 27-OCT-2000 (first entry)

DE Human endostatin protein sequence.

XX Angiogenesis-inhibiting protein receptor; angiogenesis; angiotatin;
 KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
 KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
 KW cerebral collateral; arteriovenous malformation; rheobosis; cancer;
 KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;
 KW Helicobacter related disease; fracture; cat scratch fever.

XX Homo sapiens.

OS WO200032631-A2.

PN 08-JUN-2000.

PD 06-DEC-1999; 99WO-US28897.

PF 04-DEC-1998; 98US-0206059.

PR (ENTR-) ENTREMED INC.

PA MacDonald NJ, Sim KL;

PI Macdonald NJ, Sim KL;

DR WPI; 2000-412290/35.

PT New angiogenesis-inhibiting protein receptors, useful in methods for
 PT treating diseases and processes that are mediated by angiogenesis -
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -
 XX
 XX Disclosures; Figure 3; 100pp; English.

XX This invention relates to angiogenesis-inhibiting protein receptors, and
 CC the DNA sequences encoding them. Angiogenesis is the generation of new
 CC blood vessels into a tissue, and normally occurs in wound healing,
 CC foetal and embryonal development and the formation of the corpus luteum,
 CC endometrium and placenta. Angiostatin is a protein (see AAB16450 and
 CC AAA68202) involved in angiogenesis, and has an amino acid sequence
 CC similar to that of a plasminogen fragment (see murine plasminogen
 CC AAB16490). Angiostatin has the ability to inhibit angiogenesis.
 CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and

CC AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein
 CC sequences of human laminin. Laminin is an angiotatin binding protein,
 CC and some of the peptides of the invention share homology with regions of
 CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
 CC angiogenesis-inhibiting protein receptor fragments of the invention. The
 CC peptides bind either angiotatin or endostatin and can be used in methods
 CC for treating diseases and processes that are mediated by angiogenesis,
 CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
 CC Crohn's disease, cerebral collaterals, arteriovenous malformations,
 CC rheobosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
 CC Helicobacter related diseases, fractures, placentaion and cat scratch
 CC fever. They are useful for the detection and prognosis of cancer. DNA
 CC sequences A628204-A628241 encode the peptides of the invention.
 XX
 SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5,7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIYNLKDLELFPSEALFSGSEGLPKGARIFSFDDKDVLRHPTWQKSYWHS 60
 DB 64 adraavpivnlkdellfpsealifsgseglpkgarifsfddgkdvlrhptwqkswbgs 123

OY 61 DPN 63
 DB 124 dpn 126

RESULT 9

AA90771
 ID AA90771 standard; Protein; 183 AA.

AC AA90771;

DT 22-AUG-2000 (first entry)

DE Human angiogenesis inhibiting factor 1 protein.

XX Human angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
 KW Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
 KW abnormal vessel disease.

XX Homo sapiens.

OS CN1244536-A.

PN 16-FEB-2000.

PD 10-AUG-1998; 98CN-0117150.

PF 10-AUG-1998; 98CN-0117150.

PR (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.

PA Yang Z, Guo W;

PI Yang Z, Guo W;

DR WPI; 2000-388168/34.

DR N-PSDB; AAA29884.

PT Angiogenesis inhibiting factor 1 and its derivative useful for treating
 PT tumors -
 XX
 XX Claim 1; Fig 5; 41pp; Chinese.

XX The present sequence represents an angiogenesis inhibiting factor (1),
 CC designated IAF-1. The present invention also describes: (1) preparation
 CC of (1) and its derivative; (2) an IAF binding acceptor and its
 CC preparation; and (3) an IAF antibody. (1) is useful for preparing new
 CC biological preparations for effectively treating various tumours and
 CC abnormal vessel diseases. The IAF antibody is preferably a polyclonal
 CC antibody, mosaic antibody, single stranded antibody and human originated
 CC antibody.

XX Sequence 183 AA;
 SQ
 Query Match 100.0%; Score 346; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLDELFPSEALFSGSECPKPGARIFSFQDKDVLRHPTWQKSWMGS 60
 |||
 DB 64 adraavpivnlkdellfpsealfsgseglkpgarifsfdgkdvlrhptwqkswmgs 123
 OY 61 DPN 63
 |||
 DB 124 dpn 126

RESULT 10
 ID AAY70252 standard; Protein: 183 AA.
 *XX AAY70252;
 AC
 XX
 XX 06-JUN-2000 (first entry)
 DT
 XX
 DE Human angiogenesis inhibitor, endostatin.
 XX
 XX Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
 KM angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
 KM antiproliferic; antidiabetic; ophthalmological; immunosuppressant;
 KM vasoregic; vulnerary; treatment; antiarteriosclerosis; tumor;
 KM metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KM ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KM myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KM wound granulation; keloid scar; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200011033-A2.
 XX
 PD 02-MAR-2000.
 XX
 PE 25-AUG-1999; 99WO-US19329.
 XX
 XX 25-AUG-1998; 98US-0097883.
 *PR
 XX (LEXI-) LEXINGEN PHARM CORP.
 PA
 XX
 PI Lo K, Li Y, Gillies SD;
 XX
 DR WPI: 2000-237616/20.
 DR N-PSDB: AAZ51291.
 XX
 PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin
 PT Fc region, useful for treating conditions mediated by angiogenesis,
 PT such as rheumatoid arthritis, tumors and macular degeneration -
 XX
 XX Example 1; Pages 41-42; 68pp; English.

The patent discloses a DNA molecule encoding a fusion protein comprising
 a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 inhibitor selected from angiotensin, endostatin, a plasminogen fragment
 having angiotensin activity, a collagen XVIII fragment having endostatin
 activity, or combinations of them. The fusion protein (immunofusin) is
 used to inhibit angiogenesis and to treat diseases or conditions mediated
 by angiogenesis. Conditions that may be treated include solid tumors,
 blood born tumors, tumor metastasis, benign tumors including
 hemangiomas, acoustic neuromas, neurofibromas, trachomas and pyogenic
 granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 e.g. diabetic retinopathy, retinopathy of prematurity, macular
 degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 fibroplasia, rubecosis and Osler-Weber syndrome; myocardial angiogenesis,
 plaque neovascularisation, telangiectasia, haemophilic joints;

CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
 CC in gene therapy. The present sequence is a human endostatin used in the
 CC construction of immunofusin containing human immunoglobulin gamma
 CC (IgG) Fc fragment.
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 346; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLDELFPSEALFSGSECPKPGARIFSFQDKDVLRHPTWQKSWMGS 60
 |||
 DB 64 adraavpivnlkdellfpsealfsgseglkpgarifsfdgkdvlrhptwqkswmgs 123
 OY 61 DPN 63
 |||
 DB 124 dpn 126

RESULT 11
 ID AAB49379 standard; Protein: 183 AA.
 *XX AAB49379;
 AC
 XX
 XX 02-MAR-2001 (first entry)
 DT
 XX
 DE Human endostatin SEQ ID NO: 2.
 XX
 XX Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
 KM cancer; inflammation; angiogenesis-dependent disease.
 KM
 XX
 OS Homo sapiens.
 XX
 PN WO200067771-A1.
 XX
 PD 16-NOV-2000.
 XX
 PE 02-MAY-2000; 2000WO-US12063.
 XX
 XX 06-MAY-1999; 99US-0132907.
 *PR
 XX 14-JUL-1999; 99US-0353333.
 XX
 XX (BURN-) BURNHAM INST.
 PA
 XX
 PI Vuori K;
 XX
 DR WPI: 2001-040937/05.
 DR N-PSDB: AAC88289.
 XX
 PT Endostatin peptide comprising at least four endostatin amino acid
 PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy -
 XX
 XX Disclosure; Fig 1; 146pp; English.

The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of
 CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inappropriate invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubecosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.

SO Sequence 183 AA;

Query Match 100.0%; Score 346; DB 22; Length 183;
Best Local Similarity 100.0%; Pred. No. 5.7e-39;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELLFPSEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60
Db adraavpivnlkdelldfpsealifsgseglpkpgarilfsfdgkdvlrhptwpsvwhgs 123

OY 61 DPN 63
Db 124 dpn 126

RESULT 12
AAW90877
ID AAW90877 standard; protein: 193 AA.
-XX
AC AAW90877;
XX
DT 07-JUL-2000 (first entry)
XX
*DE Human HMW endostatin (4) protein.
XX
KM Endostatin; human; renal insufficiency; antitumor; antiproliferative;
XX treatment; angiogenesis; tumor; vascular disease.
XX
OS Homo sapiens.
XX
PN MO200017240-A1.
XX
PD 30-MAR-2000.
XX
PF 21-SEP-1999; 99WO-EP06963.
XX
PR 21-SEP-1998; 98DE-1042992.
XX 03-APR-1999; 99DE-1015267.
PR 08-JUN-1999; 99DE-1026040.
XX
PA (HAEM-) HAEWOEP PHARMA GMBH.
XX
PI Staendker L, Forssmann W;
XX
DR WPI: 2000-292826/25.
XX
PT New high molecular weight form of endostatin, useful e.g. as
-PT antiangiogenic agent for treating cancer, isolated from hemofiltrate of
PT patients with kidney failure -
XX
PS Claim 2; Page 21-22; 32pp; German.
XX
CC This invention describes novel human high molecular weight (HMW)
CC endostatin (he) protein fragments isolated from the hemofiltrate of
CC patients with renal insufficiency. The products of the invention have
CC antitumor and antiproliferative activity. he is used to treat: (i)
CC diseases that involve uncontrolled angiogenesis, particularly (1)
CC and (ii) vascular diseases of supporting or connective tissue;
CC respiratory tract, cardiovascular system, urogenital tract and nervous
CC system, or sensory organs (particularly the eye). he is also used to
CC raise specific antibodies which are used for diagnosis and treatment of
CC conditions that involve overexpression of he. he has a very long plasma
CC half-life and can be administered repeatedly without inducing an immune
CC response. AAW90874-W90877 represent the endostatin proteins described in
CC the method of the invention.
XX
SQ Sequence 193 AA;

Query Match 100.0%; Score 346; DB 21; Length 193;
Best Local Similarity 100.0%; Pred. No. 6.1e-39;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELLFPSEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60
Db 77 adraavpivnlkdelldfpsealifsgseglpkpgarilfsfdgkdvlrhptwpsvwhgs 136

OY 61 DPN 63
Db 137 dpn 139

RESULT 13
AAW90874
ID AAW90874 standard; protein: 195 AA.
XX
AC AAW90874;
XX
DT 07-JUL-2000 (first entry)
XX
DE Human HMW endostatin (1) protein.
XX
KM Endostatin; human; renal insufficiency; antitumor; antiproliferative;
XX treatment; angiogenesis; tumor; vascular disease.
XX
OS Homo sapiens.
XX
PN MO200017240-A1.
XX
PD 30-MAR-2000.
XX
PF 21-SEP-1999; 99WO-EP06963.
XX
PR 21-SEP-1998; 98DE-1042992.
XX 03-APR-1999; 99DE-1015267.
PR 08-JUN-1999; 99DE-1026040.
XX
PA (HAEM-) HAEWOEP PHARMA GMBH.
XX
PI Staendker L, Forssmann W;
XX
DR WPI: 2000-292826/25.
XX
PT New high molecular weight form of endostatin, useful e.g. as
-PT antiangiogenic agent for treating cancer, isolated from hemofiltrate of
PT patients with kidney failure -
XX
PS Claim 2; Page 18-19; 32pp; German.
XX
CC This invention describes novel human high molecular weight (HMW)
CC endostatin (he) protein fragments isolated from the hemofiltrate of
CC patients with renal insufficiency. The products of the invention have
CC antitumor and antiproliferative activity. he is used to treat: (i)
CC diseases that involve uncontrolled angiogenesis, particularly (1)
CC and (ii) vascular diseases of supporting or connective tissue;
CC respiratory tract, cardiovascular system, urogenital tract and nervous
CC system, or sensory organs (particularly the eye). he is also used to
CC raise specific antibodies which are used for diagnosis and treatment of
CC conditions that involve overexpression of he. he has a very long plasma
CC half-life and can be administered repeatedly without inducing an immune
CC response. AAW90874-W90877 represent the endostatin proteins described in
CC the method of the invention.
XX
SQ Sequence 195 AA;

Query Match 100.0%; Score 346; DB 21; Length 195;
Best Local Similarity 100.0%; Pred. No. 6.2e-39;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELLFPSEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60
Db 77 adraavpivnlkdelldfpsealifsgseglpkpgarilfsfdgkdvlrhptwpsvwhgs 136

OY 61 DPN 63

Db 137 dpn 139

RESULT 14

AAB30495
ID AAB30495 standard; protein: 216 AA.

AC AAB30495;

DT 06-MAR-2001 (first entry)

DE Amino acid sequence of vaa-endostatin fusion protein in pANT3052.

XX Streptomyces sp. strain C5. SnpA; S. venezuelae; alpha-amylase;

KM endostatin; cancer; tumour growth; angiogenesis.

XX Synthetic.

OS Streptomyces sp.

OS Homo sapiens.

*XX Key Location/Qualifiers

FT Peptide 1..28 /note= "vaa signal sequence"

FT Protein 29..216 /note= "endostatin"

WT WO200060945-A1.

PN 19-OCT-2000.

PD 12-APR-2000; 2000MO-US09747.

PR 13-APR-1999; 99US-0129084.

PA (MERI) MERCK & CO INC.

XX Desantli CL, Strohl WR;

XX WPI: 2000-686970/67.

DR N-PSDB: AAC62025.

PT Preparation of soluble recombinant endostatin involves transforming

PT Streptomyces host with expression vector comprising nucleotide

PT sequence encoding endostatin operably linked to linker and leader

PT peptide

XX Example 1; Fig 10A-B; 57pp; English.

XX The present sequence represents a fusion protein of vaa and endostatin.

CC The specification describes a method for the production of soluble,

CC recombinant human endostatin in Streptomyces. Leader sequences of

CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins

CC are linked to the N-terminal of endostatin. This ensures that endostatin

CC protein is produced as a secreted, soluble protein which needs no

CC refolding, is stable in the fermentation broth and is produced in large

CC quantities. The method is used for preparing soluble recombinant human,

CC murine or primate endostatin, which is useful in the treatment of cancer,

CC inhibition of tumour growth, inhibition of angiogenesis, isolation of

CC receptors for endostatin and for identification of anti-angiogenic

CC compounds in assays. The endostatin protein is produced as a secreted,

CC soluble protein which needs no refolding, is stable in the fermentation

CC broth and is produced in large quantities. Streptomyces are amenable

CC for cultivation in large fermentations allowing for large quantities of

CC soluble endostatin to be produced.

XX Sequence 216 AA:

Query Match 100.0%; Score 346; DB 21; Length 216;

Best Local Similarity 100.0%; Pred. No. 7, 1e-39;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAVPIVNLKDELLFSPSEALFSGSGPLKPGARIFSPGDKDYLRRPMPQKSVNHGS 60
|||||

Db 97 adraavpivnlkdeallfswsealifsgsegplkpgarilfsfsgkvdlrhpccwpqkswvhgs 156

OY 61 DPN 63
|||

Db 157 dpn 159

RESULT 15

AAW26327
ID AAW26327 standard; protein: 684 AA.

AC AAW26327;

DT 19-NOV-1997 (first entry)

DE Human alpha-1 collagen (XVIII).

XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.

XX Homo sapiens.

*XX Key Location/Qualifiers

FT Peptide 1..6 /label= GXYGX'Y' motif

FT Peptide 7..12 /label= GXYGX'Y' motif

FT Peptide 13..18 /label= GXYGX'Y' motif

FT Peptide 19..24 /label= GXYGX'Y' motif

FT Peptide 25..30 /label= GXYGX'Y' motif

FT Peptide 31..36 /label= GXYGX'Y' motif

FT Peptide 37..42 /label= GXYGX'Y' motif

FT Peptide 43..48 /label= GXYGX'Y' motif

FT Peptide 49..53 /label= GXYGX'Y' motif

FT Peptide 54..59 /label= GXYGX'Y' motif

FT Peptide 60..65 /label= GXYGX'Y' motif

FT Peptide 66..71 /label= GXYGX'Y' motif

FT Peptide 72..77 /label= GXYGX'Y' motif

FT Peptide 78..83 /label= GXYGX'Y' motif

FT Peptide 84..89 /label= GXYGX'Y' motif

FT Peptide 90..95 /label= GXYGX'Y' motif

FT Peptide 96..101 /label= GXYGX'Y' motif

FT Peptide 102..107 /label= GXYGX'Y' motif

FT Peptide 108..113 /label= GXYGX'Y' motif

FT Peptide 114..119 /label= GXYGX'Y' motif

FT Peptide 120..125 /label= GXYGX'Y' motif

FT Peptide 126..131 /label= GXYGX'Y' motif

FT Peptide 132..137 /label= GXYGX'Y' motif

FT Peptide 138..143 /label= GXYGX'Y' motif

FT Peptide 144..149 /label= GXYGX'Y' motif

FT Peptide 150..155 /label= GXYGX'Y' motif

FT Peptide 156..161 /label= GXYGX'Y' motif

FT Peptide 162..167 /label= GXYGX'Y' motif


```
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 129..134
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 135..140
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 141..146
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 147..152
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 153..158
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 159..164
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 165..170
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 171..176
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 181..186
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 187..192
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 193..198
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 215..220
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 221..226
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 227..232
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 233..238
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 239..244
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 257..262
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 263..268
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 269..274
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 275..280
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 286..291
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 292..297
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 298..303
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 309..314
FT /label- GYXGX'Y' motif
```

```
FT /note- "Claim 1"
FT Peptide
FT 315..320
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 322..328
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 329..334
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 335..340
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 354..359
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 360..365
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 366..372
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 523..528
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 542..547
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"
FT Peptide
FT 590..595
FT /label- GYXGX'Y' motif
FT /note- "Claim 1"

US5643783-A.
PN
PD 01-JUL-1997.
XX
PF 01-DEC-1993; 93US-0159784.
XX
PR 01-DEC-1993; 93US-0159784.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Oh SP, Olsen BR;
XX
DR WPI: 1997-350247/32.
XX
DR N-PDB: AAT84484.
XX
PT Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX
PS Claim 1; Column 23-30; 35pp; English.
XX
CC Novel human type alpha-1 (XVIIII) collagen is characterised by
CC 10 triple helical domains containing the GYXGX'Y' motif (where X,
CC Y, X and Y' represent any amino acid), the helical domains being
CC separated and flanked by non-triple helical regions which may
CC provide flexibility. Alpha-1 collagen is expressed in multiple
CC tissues, especially liver, lung and kidney. A claimed plasmid
CC comprising alpha-1 collagen nucleic acid (see AAT84484) and an
CC expression control sequence can be used to express recombinant
CC collagen in prokaryotic or eukaryotic (especially mammalian) host
CC cells. The alpha-1 collagen may be used to treat a patient
CC suffering from a disease associated with cartilage degradation, and
CC for supplementing collagen. It can also be used as a connective

Query Match 100.0%; Score 346; DB 18; Length 684;
Best Local Similarity 100.0%; Pred. No. 3e-38;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAPIVNLKDELFPSEALFSGSEGPLKPGARIFSPDKDVLRHPPWPKSYWHS 60
DB 565 adraapivnlkellfpsealifsgseplkpgarilifsdgkavlfnplwpqkswngs 624
```

Oy 61 DPN 63
|||
Db 625 dpn 627

Search completed: September 10, 2001, 13:49:11
Job time: 106 sec